CASA0009: Spatial Data Capture and Analysis——Group Report

Group Name: Rolling CASAers
Group members: Yiyang Dai, Jiayan Fan, Tong Li, Xinlei Yan, Yujun Yuan

Word count: 5,253 (exclude headers and tables)

Lost Children, Stolen Life

—Child Trafficking in China

1.Introduction

Children trafficking is considered one of the most serious crimes of the 20th century (Wang, 2015), Looking back at the trend of child trafficking around the world, most countries link the purpose of child trafficking with sexual slavery and labour exploitation (Martinho, Gonçalves and Matos, 2020), but considering the differences in time, space and culture, the main purpose of child trafficking does not seem to be the same in China. The phenomenon of child trafficking in China has a long history and has for a long time been a neglected crime, even considered a 'tradition' (Tan, 2018). Child abduction and trafficking is a serious and irreversible danger to the abducted child, the family of origin, the family that bought the child, and society at large (Wu, Liu and Duan, 2017). Of particular note is that, China experienced a large number of child trafficking incidents from the 1980s to the early 2000s, where the ultimate purpose of these cases is more linked to family inheritance and getting rid of the moral humiliation of being childless (Liu, 2018), so the perpetrator is both a participant in trafficking and possibly an illegal “guardian” who protects the child’s growth (Ren, 2004) . This makes child trafficking in China complex with specific characteristics and factors.

The purpose of this article is to analyse the characteristics of child trafficking in China through a visual and interactive approach. In section two, we introduce the background information and conduct literature review. Then we describe the data processing in section three . Detailed data analysis, visualisation methods and interactive integration are explained in sections four, five and six respectively, based on which we form our conclusions in section seven.

2. Background

2.1. Literature review

Most of the literature on Chinese-style child trafficking comes from within China. Previously, due to the lack of relevant datasets, most studies in China focused on the use of qualitative analysis to study individual children. For example, Wang (2015) used the theory of multi-legislative co-governance to propose that the current Criminal Law, which mainly deals with the crime of child trafficking in China,

needs to be combined with other policies, such as the Adoption Law, to ensure that couples who would been involved in trafficking and illegal adoption have formal channels to adopt a child. In addition, Fan (2019) conducts interviews and questionnaires on children in a community and finds that, for individuals, children's weak awareness of prevention is one of the key reasons for trafficking. Some qualitative analyses study the context of child trafficking. Among these, Wang (2005) discusses the ways and types of children trafficked through interviews and group interviews, and finds that the number of child trafficking in East China was decreasing and the types were increasing. However, due to the limitations in samples and qualitative methods, the generalisability and practicality of these research results are often not high (Queirós, Faria and Almeida, 2017).

With the increase of samples and the accessibility of data, some scholars have realised the shortcomings of qualitative methods and conducted quantitative research on child trafficking. Among them, some scholars studied the trend of trafficked children over time in the 20th century through a bar chart, but did not mention the factors that caused the trend to change (Tan, 2018). In addition, the article by Wu, Liu and Duan (2017) mentions that the economic and social factors can also affect the difference in the number of trafficked children, but did not discuss it in detail.

It can be seen from the above that the research on Chinese-style child trafficking has a long history. The above-mentioned articles try to uncover the characteristics of child trafficking in this aspect from five different aspects: individuals, space, process, time and factor. However, given the one-sidedness of the data and qualitative methods, and the general over-reliance on superficial narratives, they fail to explore and reveal the social problems behind China's child trafficking. This report refers to the directions involved in these literatures, and through a large amount of data analysis, using a quantitative-led visualisation research method, and combining the results with the literature, aims to obtain a characteristic of child trafficking that fits most of the situation in China, so as to fill the research gap. In addition, this paper also makes the results into an interactive web page (Appendix 1), so that the academic research can be browsed in an easy and understandable way, expanding the dissemination and influence (Cao, Wang and Yang, 2007).

2.2 Definitions

Some definitions need to be clarified before conducting analysis. As standards for child trafficking vary from country to country, in this report the Chinese government's definition of child trafficking as "abduction, acquisition, trafficking and transit for the purpose of selling" (Supreme People's Court of the People's Republic of China, 1997) is adopted. In addition, we use the Palermo Protocol, an international protocol that has been ratified by the Chinese government, as a standard definition of children, where children under the age of 18 are defined as the research object (Li et al., 2017).

3. Data introduction and processing

The main data for research is the existing public second-hand data crawled from the Chinese family tracing website "Baby Go Home" (http://www.baobeihuijia.com/), mainly including the serial number, name, date of birth, date and place of trafficking, current location of the missing person, and supplementary information, a total of 35,434 entries of data. The date of data collection ends in November

2021. Considering data privacy, the dataset is anonymized by removing the names and original serial numbers of the trafficked children.

Several steps are conducted for data cleaning. Firstly, we removed rows with unknown date of birth, trafficking date and address. Secondly, as more than half of the entries ignore city-level addresses in the description and only describe county-level addresses or have typos, they were therefore queried and manually proofread. Then, we cleaned both 'place of trafficking' and 'current location of the children' as formats of provinces and cities, for the convenience of locating cases on maps and other dataset. We deleted records whose location (either disappearance or current) do not have province and city information. Then we calculated their ages of disappearance from date of disappearance and date of birth, and filtered records within the range of 0-17, according to the definition. Finally, 26,770 clean records are left for our analysis.

In addition to the main dataset, we collect and clean data on the social causes identified during the production process that may affect the number of child trafficking, such as illiteracy rate, Gini coefficient, unemployment rate and national income. Considering that the 1990s was the peak period of child trafficking (Wu, Liu and Duan, 2017), provincial-level data from the fourth census of China in 1990 are collected (National Bureau of Statistics of People's Republic of China, 1990). After cleaning, for management, we have established a shared storage space for data and results through Github.

4. Analysis

To analyse the characteristics of child trafficking in China, the analysis is divided into five parts including "who" are most likely to be the trafficking target, "where" are children trafficked from/to, and "how", "when" and "why" are they trafficked.

4.1. Who

In this part, the article examines the average number of boys and girls trafficked by age group in each province, using data related to individual characteristics (age and gender) in the main dataset. In addition, word cloud analysis serves as an auxiliary analysis to help understand gender differences in child trafficking. Considering that many people did not know the age at which they were abducted and wrote 0 years old, which may lead to data distortion, 0 years old was not included in the analysis. In addition, because the number of word clouds after the age of 9 is greatly reduced, it is not involved in the analysis. The result is as follows:

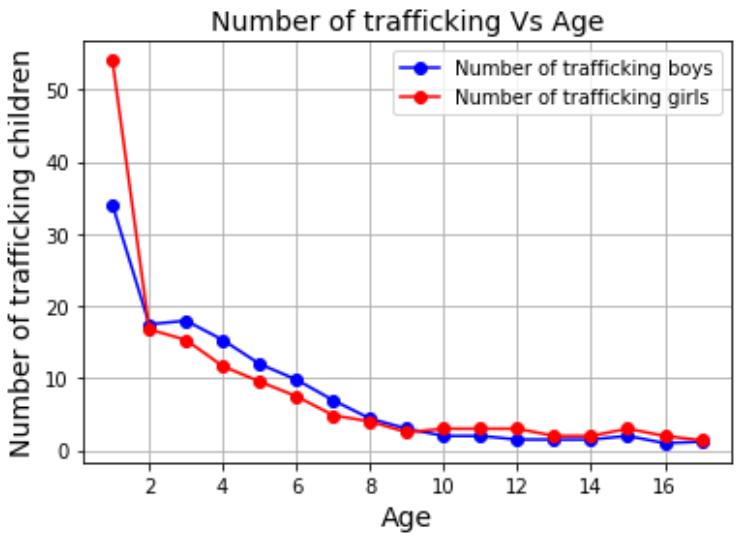

Figure 4.1.1: The provincial average number of trafficked children's gender distribution by age (1-17)

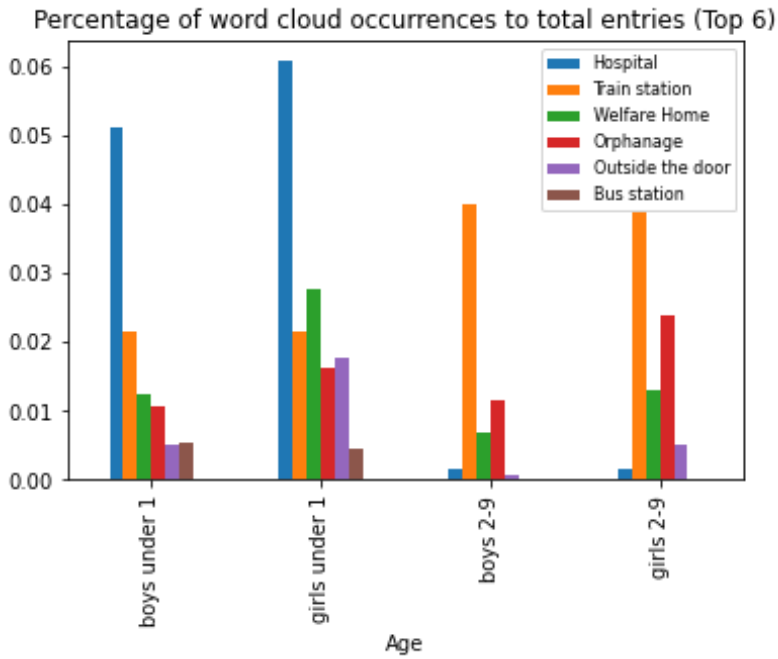

Figure 4.1.2: Percentage of word cloud occurrences to total entries (Top 6) by different age

Based on the figure 4.1.1 and figure 4.1.2, different genders show different levels of trafficking risks at different ages.

Firstly, before the age of two, there are more girls trafficking than boys'. We make two word-clouds of trafficking places feature for male and female younger than 1-year-old (including 1) separately, and find that the word "hospital" appears in twice as many female's word-clouds as male's. Hospitals are considered hard-hit areas for children abandoned and trafficked due to patriarchal ideology (Huang, 2017). This may also explain why more girls than boys go missing during this period.

For children aged 2 to 9, more boys are trafficking, while “train station” becomes the most common word in the word-clouds. Tan (2018) believes that the public space such as train stations and markets with high traffic flow are the most likely places for children to be trafficked. These children are always sold and used to pass on the family name and bring up children for the purpose of being looked after in old age. Considering the importance of boys in traditional Chinese families, this may explain why more boys disappeared during this period.

Thirdly, for the age 10 to 17, the number of missing children has dropped significantly, while the number of missing girls is more than the boys’. The reason for the significant decrease may be because with the growth of children, who are difficult to be trafficked (Tan, 2018). Moreover, during this age, the fertility of young women already exists. Xu (2013) believes that some rural people like to buy a child bride for themselves or their son to reduce the cost of marriage.

4.2. Where

To analyse the geographical distribution of child trafficking and flows, the cases are first categorised by their trafficking origins and destinations into each province or city, ranked by the number, and then presented in several layers in an interactive map (see Figure 4.2.1 and Figure 5.2.1). The distribution of numbers of trafficking cases into provinces shows that provinces in the East part of China generally have more children trafficked in, where Henan and Shandong are the top two destinations. On the city level, Putian in Fujian and Xuzhou in Jiangsu have the highest numbers. As for the trafficking origins, Jiangsu and Xuzhou are the top two provinces, and Shanghai and Chengdu are the top two cities. The distribution patterns are also seen by the density analysis of case locations which produces the heatmaps, where the centre of East China has a large cluster of trafficking cases, and three smaller clusters are seen in Guangdong, Fujian and Sichuan provinces. Differences between neighbouring cities are analysed by the Local Moran’s Index, where HH points are concentrated in the centre of East China, and several HL points are found in the west including some major cities.

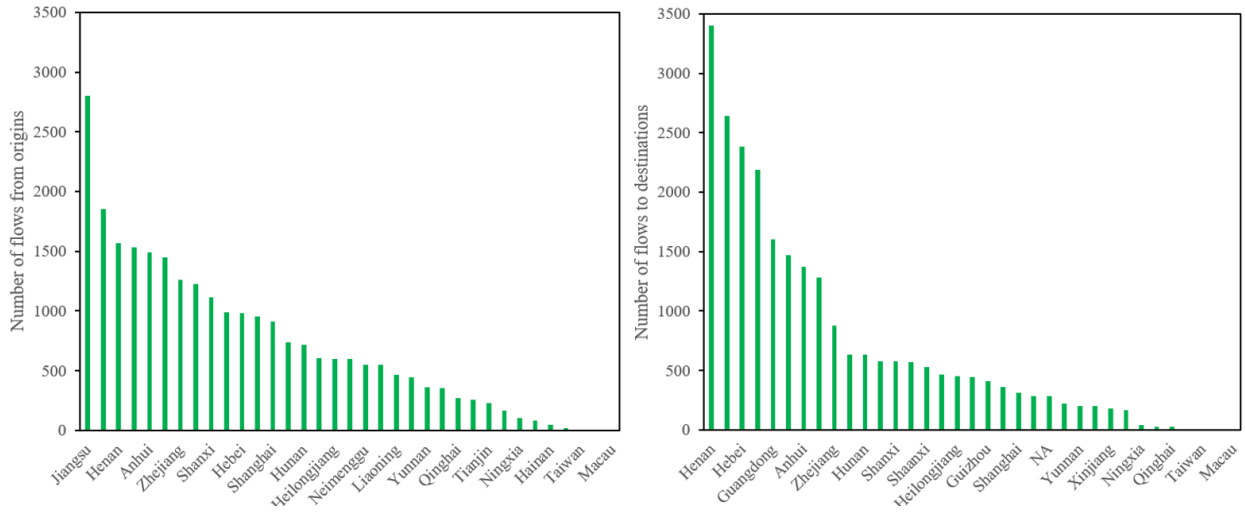

Figure 4.2.1: Statistical summary of the number of flows of each province

The trafficking flows between provinces are then summarised by each route and presented in an interactive network graph (see Figure 4.2.2 and Figure 5.2.3). It can be observed that the flows are generally from west to east, and south to centre. The top two links with the highest flows are both from Jiangsu, which is the top one origin, to Henan and Shandong, which are the top two destinations. Provinces in East China have links almost to all the other provinces. By comparing the internal flows with each province, it should be noted that although Shandong and Jiangsu remain the top two, Guangdong, Fujian and Sichuan, which are the three smaller clusters observed in the heatmaps, have large internal flows with relatively smaller external flows, which means that child trafficking between local cities within the province is serious.

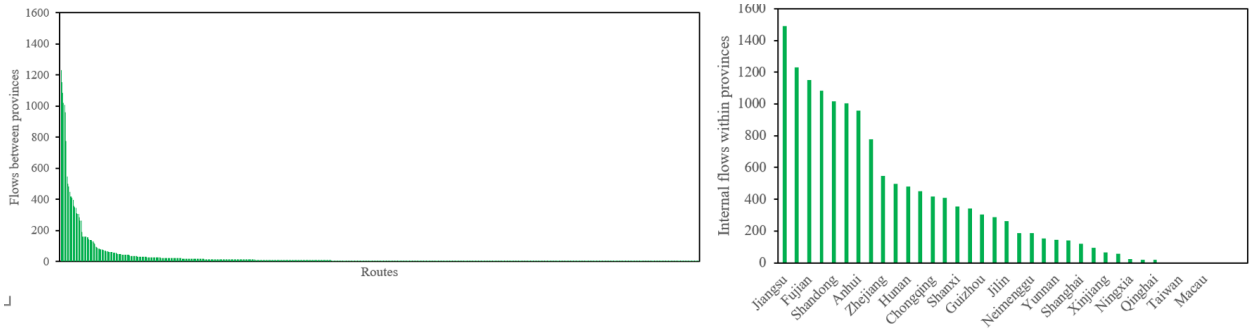

Figure 4.2.2: Statistical summary of the number of flows between provinces

To analyse the precise location of child trafficking, the locations are concluded with high-frequency words based on the word segmentation. from table 1, hospitals and train stations are the locations that highly occur the trafficked, and the explanation of the doorway is mixed with various places, including in front of the government, factories, homes, vegetable markets, public security bureaus, and more according to the text description. Since the further analysis of the relationship between the location, trafficked number, age, and gender have already been mentioned above, we will supplement some information with the text description. The above analysis shows that most of the children are under 1 year old and we can also consider where they get the information source since 'infantile amnesia' can impact the memory of infants (Ikeda, 2017). From the text description, children have the following sources including their foster parents, neighbour, relative, and more. Even though the sources may vary, we can still study them with further research.

Table 1. Location of Child Trafficking From Text Description

Rank	Location	Frequency
1	Hospitals	1744
2	Train Stations	834
3	Charities	714
4	Orphanages	516
5	Doorways	416
6	Bus Stations	164

To conclude, child trafficking is more serious in the east, where the flows are generally from west to east and south to centre. Also, a large portion of children are trafficked in hospitals and train stations.

4.3 How

The text description is then utilised to find how children are trafficked from their original families to foster families. To identify the flowing in and flowing out, we separate the analysis into two categories.

First of all, flowing in (see Table 2) includes two major ways: adoption and pickup. Adoption can also be divided into private adoption and public adoption and the difference is whether they go through the process of a state agency. Private adoption is more likely to happen in China. From the text description, the reasons behind the private adoption are the following: the proof of family's financial status can not go through the check of a state agency, afraid the one-child policy may cause the children's parents to lose their job, or their original parents pass away or are in jail. Many parents choosing to have a private adoption rather than public adoption may be more concerned about the procedure of the adoption process. Since private adoption is in the grey area to avoid the potential penalty from the one-child policy or the parents may lack legal awareness, they will consider private adoption as an easier way.

Secondly, flowing out has three channels (see Table 3) which include: abandoning, giving away, and missing. Giving away happens the most for children to flow out from their original families. From the text description, the reasons behind the giving away are the following: the baby has a serious illness, the family's financial situation can't afford to raise the children, or the original family prefers a boy to a girl. Still, many of the reasons for being given away by the original family remain unknown.

Overall, from the reasons behind the private adoption and the giving away, we can find what may cause the children to flow in and flow out. From people’s description, the major internal factors are the family’s financial situation and awareness and the major external factors are the effects of the undergoing policies which can have further research.

Table 2. How Children Flow in From Text Description:

Rank	How	Frequency
1	Private Adoption	1703
2	Public Adoption	1317
3	Picked up	521

Table 3. How Children Flow out From Text Description:

Rank	How	Frequency
1	Giving Away	1530
2	Abandoned	1377
3	Missing	533

4.4 When

4.4.1 Yearly changes of the total number of trafficked children

The preference for boys has always been widespread in China (Li, Marcos 1999), but the family planning policy implemented in the early 1980s has certainly exacerbated this phenomenon (Hansel, M. H. 2002). The traditional notion that only male heirs could continue the 'bloodline' and the reality that only one child could be born fuelled the demand for male babies and children, but also led many families to abandon, give up or even sell their daughters. In addition, during the 1980s, the increased mobility of Chinese society (Jin, Wang 2004) and the rapid development of rail transport (Ibid.) led to more and more people who once lived in the rural areas taking their families to the cities to earn a living, providing opportunities for traffickers to steal children. A large market for the trade in children was thus formed, with the number of transactions reaching a peak around 1989.

In the 1990s, the Chinese government launched a vigorous 'Daguai' (anti-trafficking campaign), and the rampant child trafficking crime began to decline year on year. But the absolute number of trafficking cases was still at a high level.

Since 2000, as the government has gradually loosened the number of births, from couples where both from a single-child family could have two children (2002), to couples where either the husband or the wife is from a single-child family could have two children (2013), to all couples could have two children (2016), the number of trafficked children cases is dropping continually (Li, et al. 2017). Besides, as socio-economic development and average education levels have increased, people have become less obsessed with having male offspring, which means the demand market is shrinking. As a result of the

combination of these factors, the number of child abduction cases fell significantly to a low level in the early 21st century.

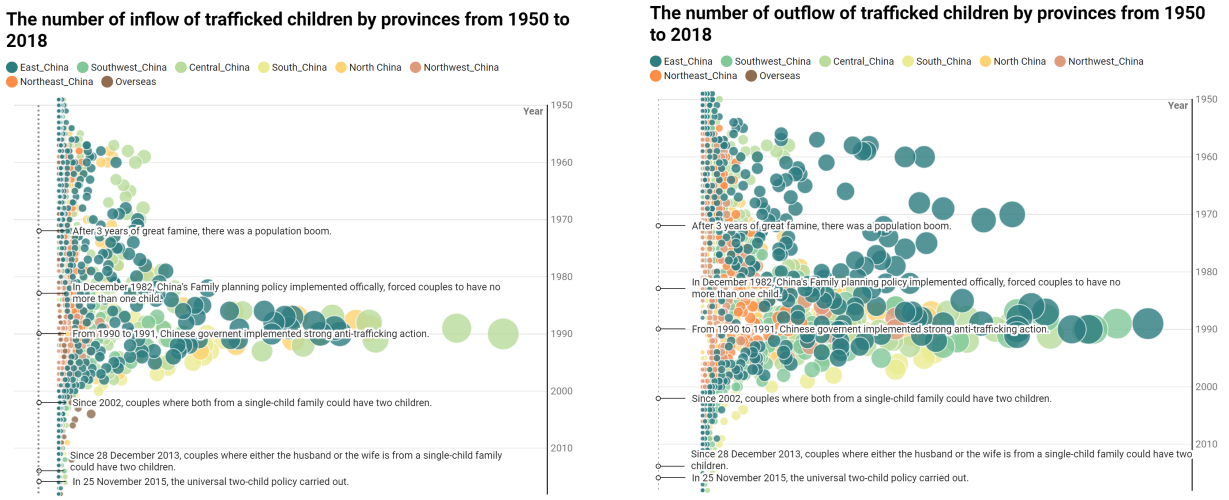

Figure 4.4.1: the number of inflow and outflow of trafficked children by province from 1950 to 2018

4.4.2 Monthly changes of the total number of trafficked children

The crime seasonal variation is low, with a more even distribution across the seasons. But January has the most child missing cases, probably due to the Spring Festival when parents are more likely to go out with their children to crowded places, and people tend to move long distances across provinces to return hometown during that time, giving the chance to traffickers to stole and transit victims.

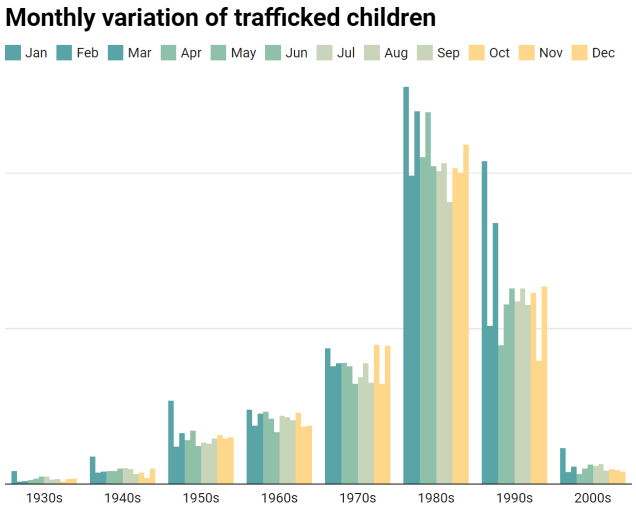

Figure 4.4.2: Monthly variation of trafficked children

4.5 Why

The results of the relationship between child trafficking and economic/social characteristics at the province level are as follows.

The first plot shows a positive relationship between wage and child trafficked in cases, which indicates that more children are trafficked into places with higher income. The second plot shows a negative relationship between gini coefficient and child trafficked in cases, indicating that more children are trafficked into places with smaller income gaps. The third plot shows a negative relationship between unemployment rate and child trafficked in cases, which indicates that more children are trafficked into places with lower unemployment rate. The fourth plot also shows a negative relationship between illiteracy rate and child trafficked in cases, indicating that more children are trafficked into places with lower illiteracy rate.

Based on the background, news report and literature of child trafficking, we assume that child trafficking is more likely to occur in places with relatively lower average wage/ higher poverty rate and higher unemployment rate (buying a child is cheaper than paying penalty when you give birth to more than one child under One-child policy), higher income gap (higher economic inequality), higher illiteracy rate (residents are bound by traditional bad practices, e.g. preference for boys). However, the results above are opposite to our hypothesis, which is confusing.

The possible explanation is the ambiguity of the data. It only contains information about Current location of missing children, so we regard it as Places children trafficked in. However, it might be very likely that Current location of them are places they work and live after they grew up and left their adopted families, but not Places they trafficked in. Therefore, we should be critical about the results we got.

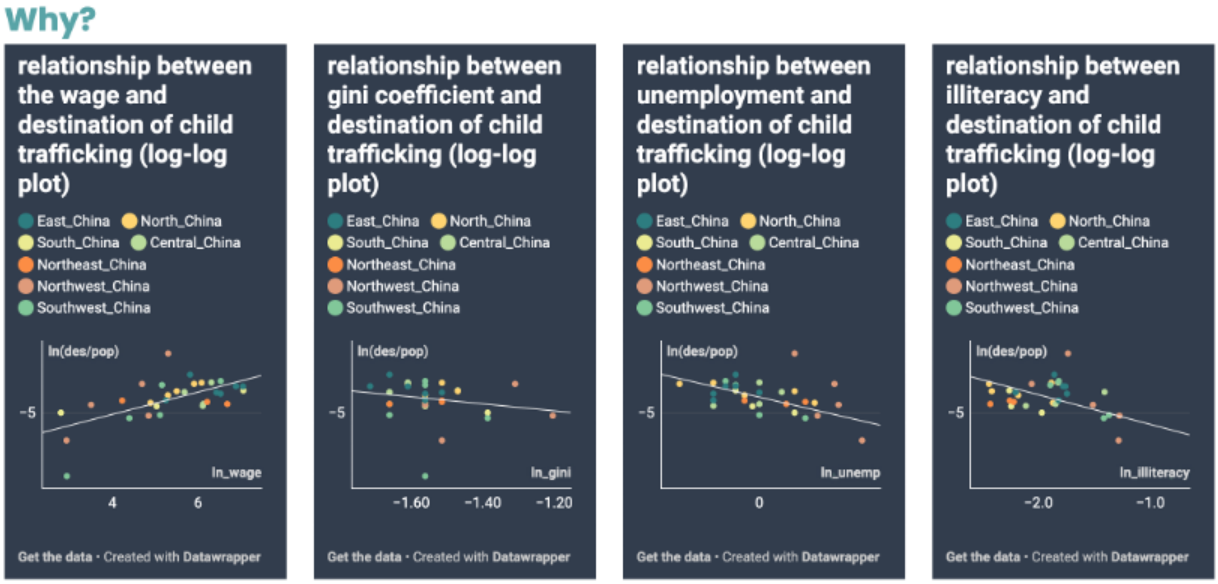

Figure 4.5.1: relationship between wage, gini coefficient, unemployment rate, illiteracy rate and child trafficking

5. Data visualisation

5.1 Statistical charts making

We use Datawrapper (www.datawrapper.de/) to create the interactive statistical charts on the website. This tool allows users to upload user's own dataset and choose a suitable type of chart to present the data. The style and colour of charts can be edited freely; extra pop-up boxes and highlight lines can be added to explain detailed information.

Take the timeline chart as an example. The first step is uploading the processed data. We picked target columns from the original dataset (in the timeline chart, the most important item we concerned is Year) and manipulated them into the suitable format in Python. After uploading, Datawrapper will classify columns according to their types and properties for subsequent visualisation.

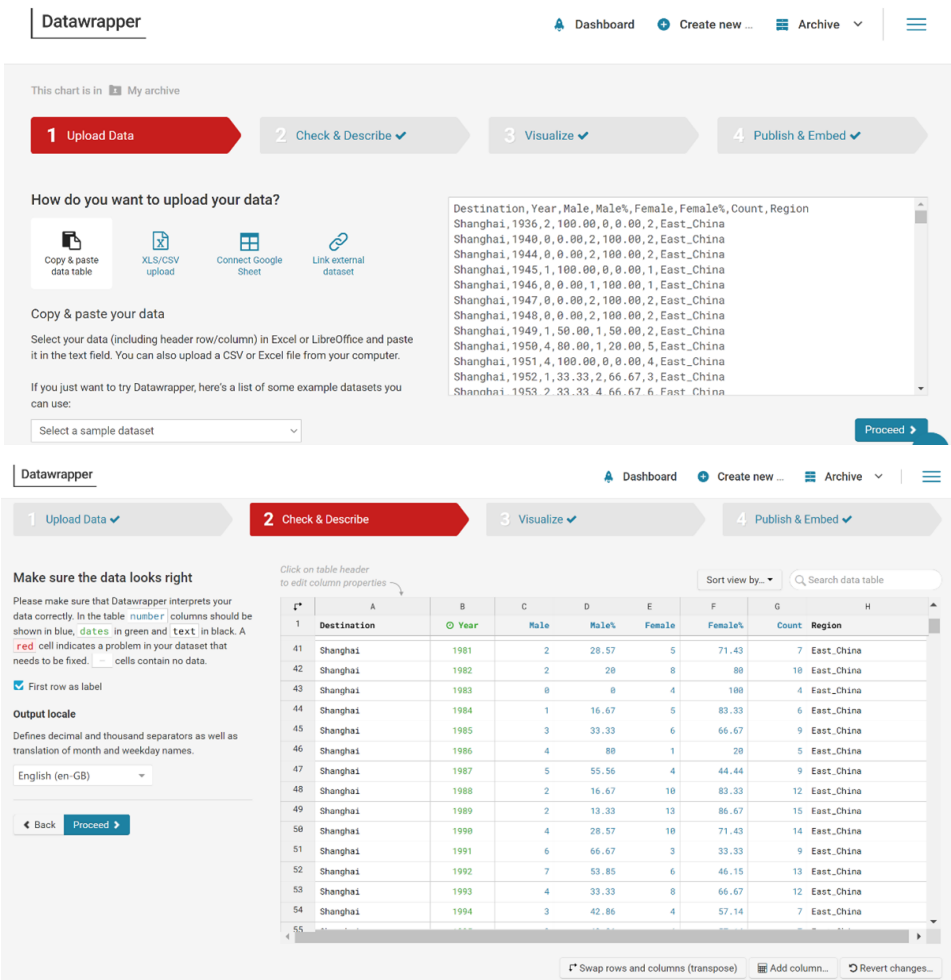

Figure 5.1.1: Statistic charts visualising step 1 and 2 in Datawrapper

In the visualising part, Datawrapper provides multiple types of charts. Due to the large number of rows for yearly data, and each item contains several values, to retain and present all data as much as possible, the scatter plot was chosen.

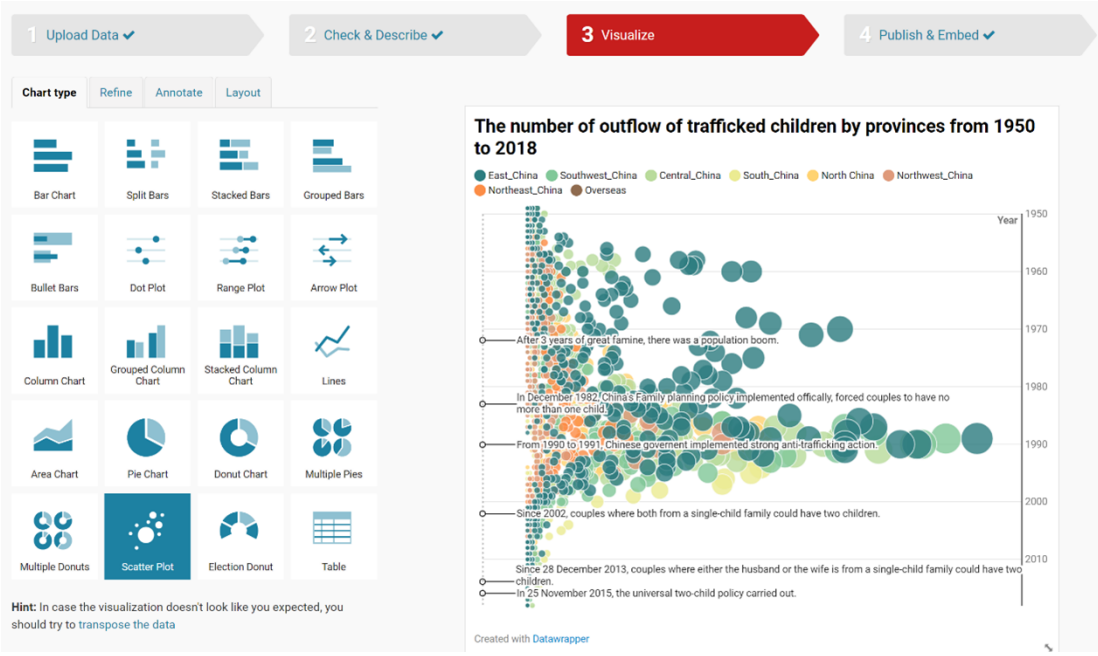

Figure 5.1.2: Statistic charts visualising step 3 in Datawrapper

The most important settings are made under the refine and annotation tabs. Refine tab allows us to choose the specific columns as the Y and X axis, and the colour of dots (case number of one province in a specific year) according to their categories (different colours represent the region of the province). The size of the dot indicates its quantity.

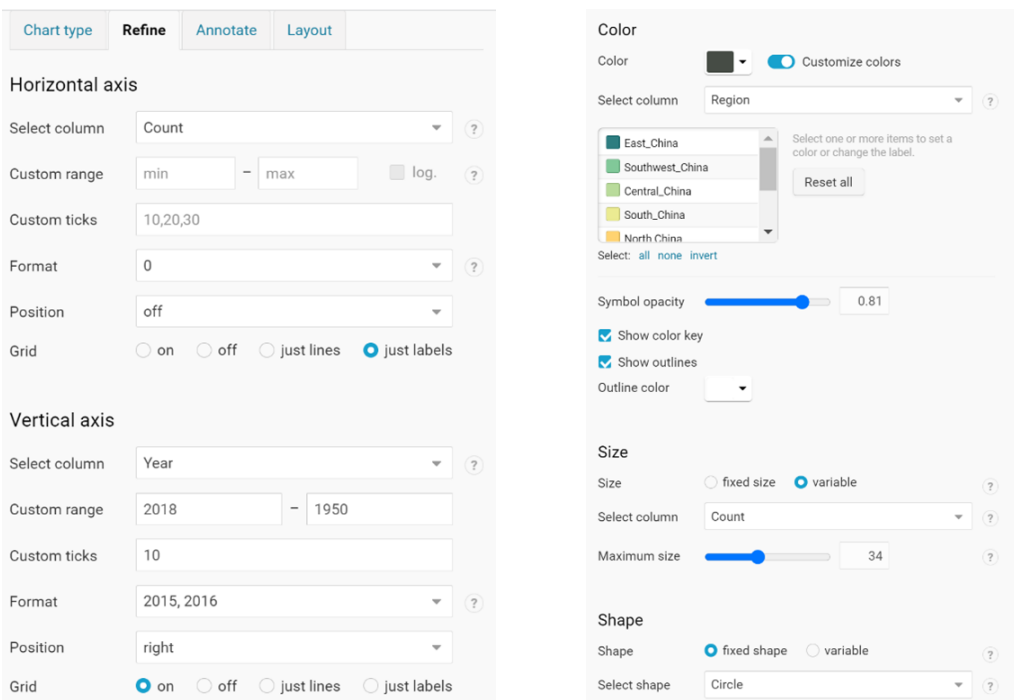

Figure 5.1.3: Statistic charts visualising step 4 in Datawrapper

Annotation tab allows to add text in a certain time point describing major events and policy. Besides, using HTML syntax in tooltip input field can add many customised effects. Here we added pop-up boxes in each dot, when users hover the mouse in one dot, the number of case and the gender ratio of victim children will show up.

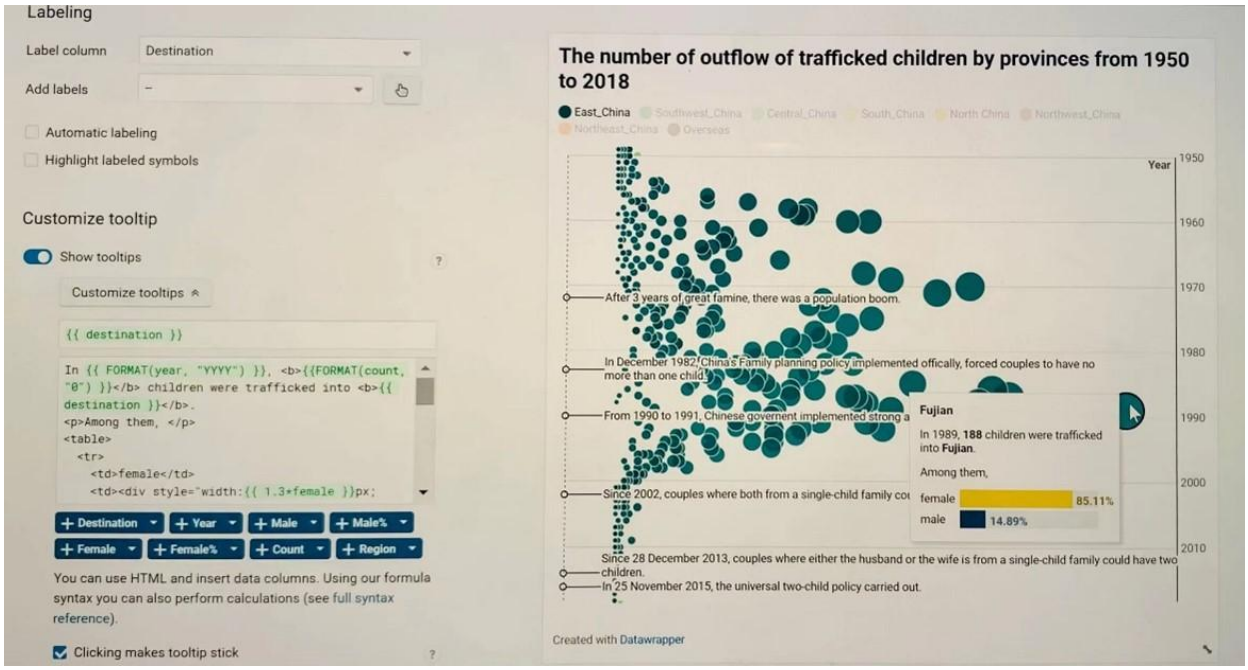

Figure 5.1.4: Statistic charts visualising step 5 in Datawrapper

In the end, Datawrapper provides embedded code which allows us to add this chart into our website.

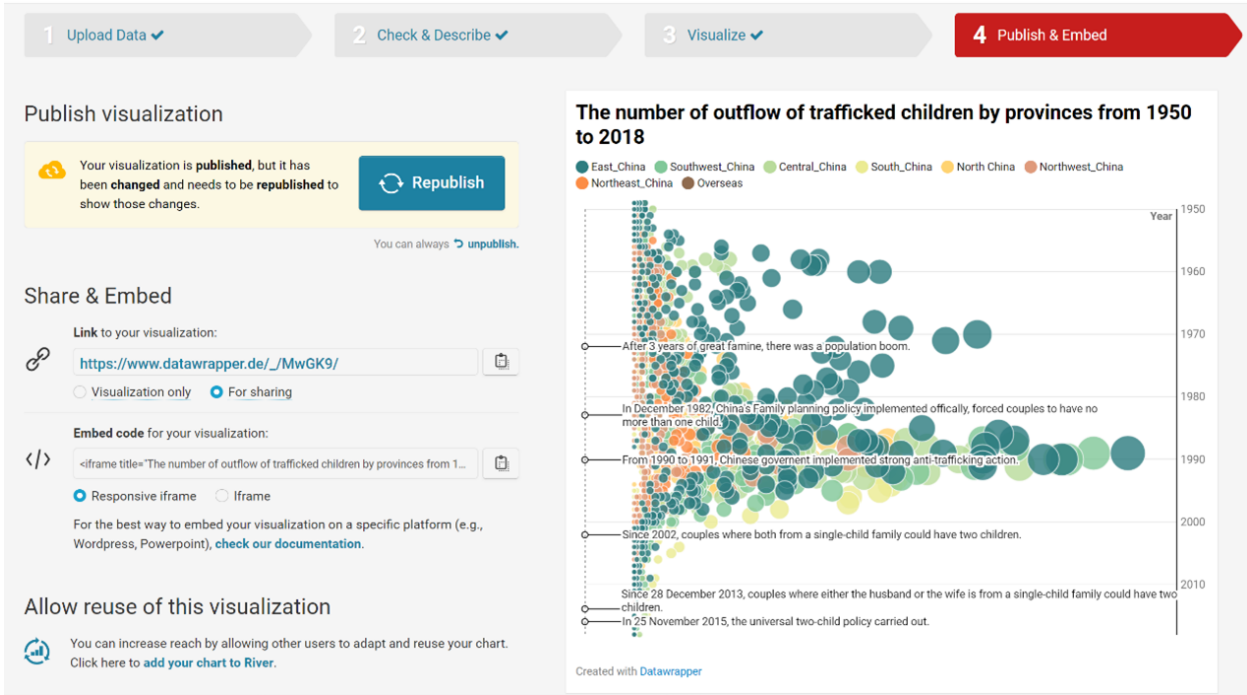

Figure 5.1.5: Statistic charts visualising step 6 in Datawrapper

By using this tool, we are able to effectively combine the main data features with external data with a certain level of interactivity and aesthetics with limited time and skill. The downside, however, is that the tool has fewer customisable sections and does not provide source code for the charts, which results in many additional desired rendering effects not being achieved.

5.2. Interactive map making

To analyse the geographical properties of the dataset from multiple aspects, interactive maps with multiple layers containing different topics are used for analysis, where two JavaScript libraries are used.

Mapbox GL JS (https://docs.mapbox.com/mapbox-gl-js/) is used to make the multi-layer choropleth map representing the trafficking case number. Mapbox GL JS defines maps that are dynamically rendered by combining vector tiles with style rules in the browser, allowing data to be displayed in response to user interaction. The map style is first designed and customised in Mapbox Studio, with the base layer tilesets added. Then the Mapbox JavaScript and CSS files are added in the head of the HTML file. The Map object is then created in the script using Mapbox API and access token, with properties such as style, zoom extent and boundary defined. Using the addSource and addLayer function, the 9 processed vector tilesets are added to the map, with their symbology, extent and opacity defined. The original status is set in the map.on('load') event. Then interactive effects are added using different functions. To switch the visible layer from the province layer to the city layer when zooming in, the zoom extent for the two layers are set where the province layer is visible when zoom extent is less than 4, and the city layer is visible when greater than 4 (see Figure 5.2.1). The toggleableLayerID object is created for the layers, allowing their visibility to change when clicking the button, refreshed with the map.on('idle') event (see Figure 5.2.2). The last step is adding the map.on('mouse') event to return the responding value of the active layer when hovering above the polygon unit (Figure 5.2.1). The complete map is created in an individual HTML file and then embedded into the index.html file.

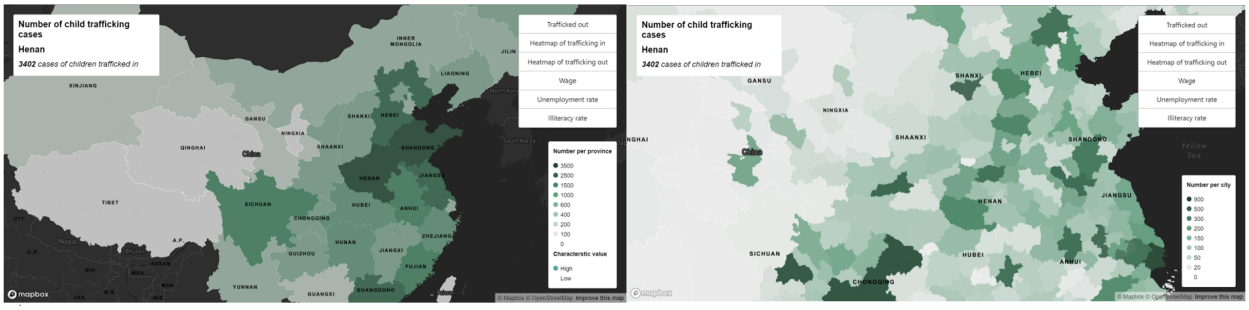

Figure 5.2.1: The zooming and mouse-hovering effects of the map

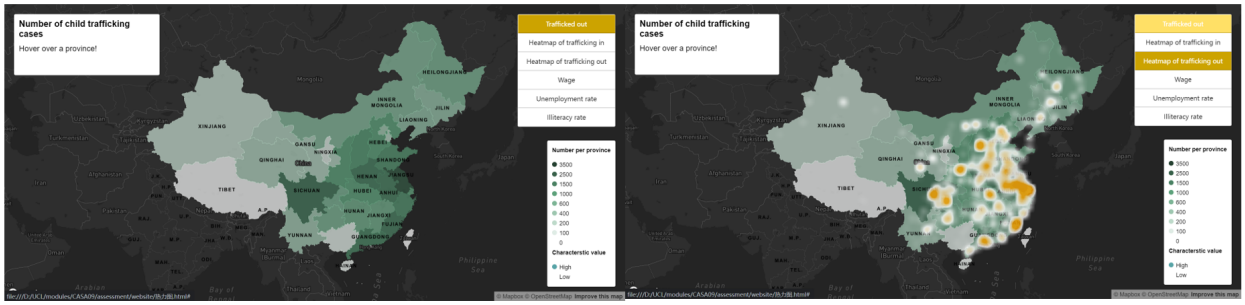

Figure 5.2.2: The multiple layers effects of the map

Apache Echarts JS (https://echarts.apache.org/en/index.html) is used to create the network graph of child trafficking flows between provinces. The echarts.js library contains various dynamic chart types for web visualisation, of which the graph series can be used to make network graphs visualising relationships. The cleaned dataset entries are first grouped into each route and put into a list of links, and the internal flows within each province are summarised into a list of nodes. The nodes and links are then put into the JSON file, where a node denotes each province, whose properties include its name, coordinate and value (number of internal flows), and a link denotes the directed trafficking route between two provinces, whose properties includes its source, target and value (number of flows). To present the graph on the web page, the echarts.js is first added in the head. The JSON file is read remotely and a graph series is created, initialising the properties of its nodes, links, and categories. The geo coordinate system and map extent are added. Using the emphasise option, the nodes and links are highlighted with their value popped when hovering over (see Figure 5.2.3). Finally, the complete map HTML file is also embedded into the index.html file.

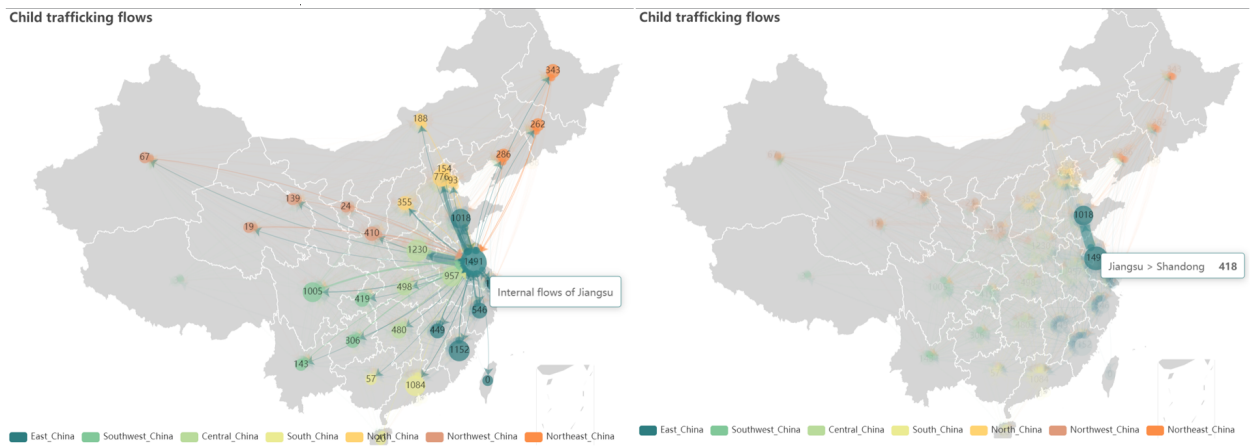

Figure 5.2.3: The hovering effects of nodes and links.

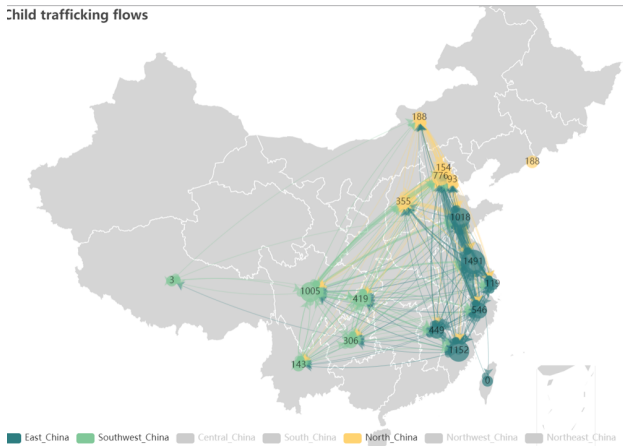

Figure 5.2.4: The grouping effects where interaction between groups can be presented separately

Moreover, for the clustering and the local Moran's I map, we utilise the Folium package in Python to draw the interactive map and then save it directly with the HTML file. The clustering result is based on the number of cases and uses MinMaxScaler to normalise. Then, we utilise K-means and calculate the k value through the Elbow method and find that 4 can be the optimal value since it starts to become stable and moves nearly parallel to the X-axis (Figure 5.2.5). For the local Moran's I map, divided into five sections based on the idea of the Lisa clusters, we first calculate the local Moran's I by the esda.moran package to classific all the points to HH, LL, HL, and LH. Then we verify the p-value to check whether each point achieves significance which is under 0.05 and the points with a higher than 0.05 p-value will be identified as non-significant.

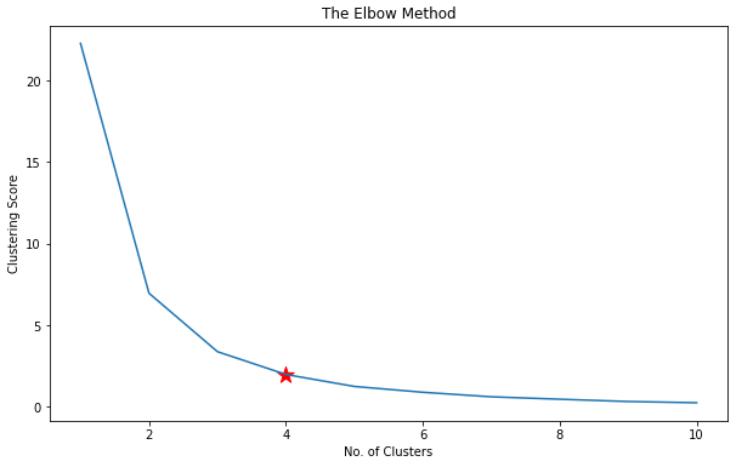

Figure 5.2.5: The elbow method to find the optimal k to the K-means clustering

5.3 Word cloud making

In the beginning, we utilise the package of the jieba in Python to segment the Chinese sentences of the description. We use this module to cut the sentences into words and then count the frequency of each word so that we can then extract the top 100 high-frequency words. Also, during this process, we consider the stop words used in the sentences and we remove the stop words to avoid the potential distraction in the analysis.

Then, we use the wordCloud package to draw the top 100 high-frequency words. However, the picture created is static and cannot be interacted with as well as 100 words will hard for readers to find the points. So we select essential words from the top 100 high-frequency words and then translate them into English. Some of the words in Chinese are meaning the same in English so we combine them as one word which may cause the words in the current wordcloud to be limited. After the translation, we start to use the echarts-wordcloud to achieve the interactive map and it is based on the wordcloud2.js for Apache ECharts. By importing the JS file from the echarts-wordcloud in the HTML file, we can make the interactive wordcloud (Figure 5.3.1). The word will change to emphasise colour and show the value which presents the frequency of appearance.

Figure 5.3.1: The word cloud to illustrate the value of each flow in and flow out channels

6. Technical integration between elements

6.1 Integration of elements

Our website uses Bootstrap as a framework, which provides style definitions for all HTML elements (including HTMLs, CSS & JS Library). Basically, the page is divided into five sections, which are website cover, data introduction, analysis, conclusion and team introduction as the footer.

After building the overall framework of the page, we fill in the sections with the appropriate text, maps and charts, and use bootstrap's grid system to lay out the elements and adjust their height and width. To ensure overall stylistic and aesthetic harmonisation, we set a colour scheme with green as the dominant colour and applied it in style.css. The external charts referenced are also edited using a consistent colour scheme. In addition, the fonts, margins, alignment of elements and animation effects are all standardised in this file.

6.2 Web interaction design

To make it easier for users to interact with the web page and to read information more intuitively, we used multiple JS libraries to achieve interactive effects on the page. For example, swiper (swiperjs.com) helps us to add a touch slider and implement the slideshow effect in the page, which is particularly suitable for displaying a topic that contains multiple charts or images.

In addition to the practical interaction, we also wanted to implement some interactive effects to make our content more compelling, especially given the heavy topic of our research, where each individual case means the breakdown of a family and the painful experience of the person concerned, which lasts for many years. Instead of starting with the background and body of the article, we first quote the parents of many famous child abduction cases in the cover section and placed the corresponding official news or Wikipedia links (in green), as well as the self-descriptions of the abducted children in our dataset (in orange), in the hope that this presentation would evoke a sense of shock among users. The hope is that this presentation will evoke in users a sense of shock at this terrible crime and sympathy for the victims of child abduction, and that they will read the rest of the book with such feelings.

The graphics and web pages are as good as we expected them to be at the time of the pitch, but due to limitations of time, skill and experience, there is still much to improve. For instance, the navigation bar could appear after the research background of the cover section to ensure the integrity of the page cover as much as possible, the combination of charts, maps and text is a bit stiff and could be improved aesthetically.

7. Conclusion, limitation and further work

7.1 Conclusion

In conclusion, firstly, for the WHO part, more girls were missing in age before 2 and 10-17, and more boys were missing in age 2-9. Secondly, for the WHERE part, child trafficking flows are generally from west to east and south to centre. And many children are found in hospitals and train stations. Thirdly, for

the WHEN part, the 1988-1990 period and January are the years and months of the peak of child trafficking. Fourthly, for the HOW part, ‘giving away’ and ‘private adoption’ are the most common situations after a child being trafficked. Lastly, for the WHY part, the reasons behind the child trafficking phenomenon can be explained by different economic and social factors. The fundamental reason behind the phenomenon may also be based on the patriarchal thought in traditional culture. In this idea, both women and men are victims.

7.2 Limitations

In this project, we define five major respects in who, where, how, when, and where in order to find significant patterns in the trafficking of children. However, there still exist several limitations in two different aspects.

There are three limitations in the data. The description data is messy, containing a lot of information that requires data cleaning. Potential inaccuracy due to the information based on memory or hearsay rather than facts may affect the results. The lack of supplementary datasets can also lead to inaccuracy, which can be replicated with better quality or updated supplementary data.

There are four limitations in the analysis. Based on the limitations of the data, there are cases neglected such as children at the age of 0 which are not further analysed. The cultural context of geographical segmentation is not considered. For instance, Xuzhou is at the border of three provinces and the movement of the population will cause cultural integration and affect trafficking cases. The different wording may impact our text mining process and cause misunderstanding in the word search or translation. Different standards of the provinces when counting the number may cause the different analysis results.

7.3 Further work

This project can be improved further in the following four areas. For the dataset, the data quality can be improved by further manipulation, like using Natural Language Processing to identify the description of each case, and make more detailed classifications. And for the data quantity, by finding proper solutions to explain the removed value, there can retain more items and reduce bias as much as possible. For analysis, it is necessary to focus on smaller regional levels (like city and county levels), especially for those main inflow and outflow cities and counties. Using the research result and combining extra data, some predictions can be proposed. For example, due to the family planning policy, there is a serious imbalance in gender in China, and women may be the main target for population trafficking in the future. Since the geospatial distribution of child abduction cases is already known, if only overlaying it with the gender ratios of births in each city in recent years, we will arrive at areas where the demand for human abduction is high, and the number of males is abnormally high, which would possibly the potential destinations for trafficked women.

Reference

Cao, K., Wang, H. and Yang, Y. (2007). ‘Improve high-traffic interactive website performance’. *Computer Engineering and Applications*, 43 (12), pp. 170–172.

Fan, P. (2019). *Practical research on group work intervening in the development of migrant children’s anti-abduction awareness*. master thesis. Gansu University of Political Science and Law. Available at: https://cdmd.cnki.com.cn/Article/CDMD-11406-1019076951.htm (Accessed: 14 May 2022).

Hansel, M. H. (2002) China’s one-child policy’s effects on women and the paradox of persecution and trafficking. Southern California review of law and women’s studies. 11 (2), 369–.

Huang, D. (2017). ‘Research on the Legal Regulation of Abandoned Children in Hospitals’. *Journal of Shanghai Academy of Social Sciences*, 2017. Available at: https://wap.cnki.net/touch/web/Dissertation/Article/87903-1017122654.nh.html (Accessed: 14 May 2022).

Jin Fengjun, Wang Jiao'e. (2004). The expansion of China's railway network and its spatial accessibility in the 20th century[J]. Journal of Geography, 2004, 59(2): 293-302.

Li, G., Li, H. and Bai, X. (2020). ‘Analysis of temporal and spatial characteristics and influencing factors of child trafficking crimes in China based on GIS’. *Journal of Henan University of Technology*, 23 (3), pp. 56–67.

Li, G., Wang, H., Tan, R. and Fu, Y. (2017). ‘Spatiotemporal characteristics and formation mechanism of child trafficking in China based on" successful cases"’. *Geography Research*. (Crime Geography), 36 (17), pp. 2505–2520. doi: 10.11821/dlyj201712019.

Liu, Y. (2018). *Research on Chinese ‘Trafficking of Women and Children’ Movies*. Hunan Normal University. Available at: https://cdmd.cnki.com.cn/Article/CDMD-10542-1018149254.htm (Accessed: 16 May 2022).

Martinho, G., Gonçalves, M. and Matos, M. (2020). ‘Child trafficking, comprehensive needs and professional practices: A systematic review’. *Children and Youth Services Review*, 119, p. 105674. doi: 10.1016/j.childyouth.2020.105674.

National Bureau of Statistics of People’s Republic of China. (1990). *Main data of China’s fourth census*. *Main data of China’s fourth census*. Available at: https://www.yearbookchina.com/naviBooklist-N2013090221-1.html (Accessed: 3 May 2022).

Queirós, A., Faria, D. and Almeida, F. (2017). ‘STRENGTHS AND LIMITATIONS OF QUALITATIVE AND QUANTITATIVE RESEARCH METHODS’. *European Journal of Education Studies*, (0). doi: 10.46827/ejes.v0i0.1017.

Ren, X. (2004). *Trafficking in Children: China and Asian Perspective*. California State University. Available at: https://scholar.googleusercontent.com/scholar?q=cache:ycoLw2brq2UJ:scholar.google.com/+Asian+child+trafficking&hl=en&as_sdt=0,5 (Accessed: 4 May 2022).

Supreme People's Court of the People's Republic of China. (1997). *Criminal Law of the People's Republic of China*. Available at: http://gongbao.court.gov.cn/Details/f8e30d0689b23f57bfc782d21035c3.html?sw=%E4%B8%AD%E5%8D%8E%E4%BA%BA%E6%B0%91%E5%85%B1%E5%92%8C%E5%9B%BD%E5%88%91%E6%B3%95 (Accessed: 5 May 2022).

Tan, R. (2018). 'Research on the Crime of Child Abduction and Trafficking in China from the Perspective of Geography'. *Journal of Xi'an University*, 2018. Available at: https://wap.cnki.net/touch/web/Dissertation/Article/10697-1018108976.nh.html (Accessed: 14 May 2022).

Wang, J. (2005). 'A case study on the inflow areas of abducted and trafficked underage women and children from five East China provinces, Yunnan and Guangxi'. *Zhejiang Academic Journal*, 2005 (4), pp. 200–215.

Wang, X. (2015). 'The Current Situation and Countermeasures of Child Abduction Crime——An Empirical Study Taking F Province as an Example'. *Journal of the People's Public Security University of China*, Social Science Edition (5), pp. 21–31.

Wu, D., Liu, T. and Duan, X. (2017). 'Spatial distribution analysis of child trafficking and its driving factors'. *Remote Sensing Information*, 2017 (1). Available at: http://www.cqvip.com/qk/92871x/201701/671396287.html (Accessed: 13 May 2022).

Xu, B. (2013). *Child bride marriage research*. Zhejiang University. Available at: https://cdmd.cnki.com.cn/Article/CDMD-10335-1014114373.htm (Accessed: 9 May 2022).

Zhao, X. and Lv, S. (2010). 'On the Causes of the Crime of Human Trafficking and Its Countermeasures'. *Journal of Shandong Police College*, 2010 (2), pp. 101–105.

Ikeda, K. (2017). *Why you can't remember being a baby*. [online] qbi.uq.edu.au. Available at: https://qbi.uq.edu.au/brain/learning-memory/why-you-cant-remember-being-baby.

Appendix 1

Our web page can be viewed in following address:

Website link: http://dev.spatialdatacapture.org/~ucfnlic/rollingcasaer.github.io/index.html
Github link: https://rollingcasaer.github.io/
Github data: https://github.com/rollingcasaer/rollingcasaer.github.io