Exploring the Relationship between the Spatial Distribution of Different Age Populations and Points of Interest (POI) in China

Yiyi Huang 1,2, Tao Lin 1,3,*, Guoqin Zhang 1,3 #, Wei Zhu 1,3, Nicholas A. S. Hamm 4 #, Yuqin Liu 1,3, Junmao Zhang 1,3 # and Xia Yao 1,3

1 Key Laboratory of Urban Environment and Health, Institute of Urban Environment Chinese Academy of Sciences, Xiamen 361021, China; yyhuang@iue.ac.cn (Y.H.); gzqzhang@iue.ac.cn (G.Z.); wzhu@iue.ac.cn (W.Z.); yqliu@iue.ac.cn (Y.L.); jmzhang@iue.ac.cn (J.Z.); xyao@iue.ac.cn (X.Y.)
2 Coastal and Ocean Management Institute, Xiamen University, Xiamen 361102, China
3 University of Chinese Academy of Sciences, Beijing 100049, China
4 School of Geographical Sciences, University of Nottingham, Ningbo 315100, China; nicholas.hamm@nottingham.edu.cn
* Correspondence: tlin@iue.ac.cn

Abstract: Population spatialization data is crucial to conducting scientific studies of coupled human–environment systems. Although significant progress has been made in population spatialization, the spatialization of different age populations is still weak. POI data with rich information have great potential to simulate the spatial distribution of different age populations, but the relationship between spatial distributions of POI and different age populations is still unclear, and whether it can be used as an auxiliary variable for the different age population spatialization remains to be explored. Therefore, this study collected and sorted out the number of different age populations and POIs in 2846 county-level administrative units of the Chinese mainland in 2010, divided the research data by region and city size, and explored the relationship between the different age populations and POIs. We found that there is a complex relationship between POI and different age populations. Firstly, there are positive, moderate-to-strong linear correlations between POI and population indicators. Secondly, POI has a different explanatory power for different age populations, and it has a higher explanatory power for the young and middle-aged population than the child and old population. Thirdly, the explanatory power of POI to different age populations is positively correlated with the urban economic development level. Finally, a small number of a certain kinds of POIs can be used to effectively simulate the spatial distributions of different age populations, which can improve the efficiency of obtaining spatialization data of different age populations and greatly save on costs. The study can provide data support for the precise spatialization of different age populations and inspire the spatialization of the other population attributes by POI in the future.

Keywords: points of interest; different age populations; spatial distribution; relationship; geodetector; China; county-level administrative units

1. Introduction

Population data are widely used in geographical research. Population spatialization refers to modeling and mapping fine spatial resolution population data. It is an extremely valuable task. The traditional collection method of population data is the census-based method. Census-based methods based on administrative divisions are costly, time-consuming, have a long sampling period, and the data does not reflect the detailed spatial patterns of the population within the administrative divisions [1]. Therefore, there are some shortcomings when using census data, such as inconsistent administrative units and natural units, low temporal and spatial resolution, and lack of suitability for spatial analysis [2–4]. Population spatialization data can resolve the above problem nicely, which is suitable for combining with other geo-referenced environmental factors for analysis [5,6], and it is...
crucial for conducting interdisciplinary studies—such as disaster risk assessment [7,8], urban planning [9–11], emergency management [12] and public health [13,14]—of coupled human–environment systems [1,15].

Various approaches and auxiliary variables have been developed and applied to population spatialization during the past few decades. Common methods of population spatialization include areal weighting interpolation [16], pycnophylactic interpolation [17], dasymetric mapping [18,19], intelligent interpolation [20], etc. [15]. Researchers also continue to apply various auxiliary variables to population spatialization, such as land use [21,22], nighttime light (includes the dataset from the US Air Force Defense Meteorological Satellite Program Operational Linescan System (DMSP/OLS) [23] and the National Polar-Orbiting Partnership’s Visible Infrared Imaging Radiometer Suite (NPP/VIIRS)) [24]), as well as geographic information data (e.g., topographic features [1], residential areas [25], road networks [15], and river streams [26], etc.). New auxiliary variables such as laser radar [27,28], Twitter [29,30], mobile phone signaling [31–33], and points of interest (POI) [1,10,15] are also applied to improve the temporal and spatial resolution and accuracy of population spatialization. POI is typical geospatial big data and contains geographic coordinates and information representing different functions. Certain POI types are highly associated with human activities, indicating the population density around them [1,34]. Therefore, POI data have been applied to define the urban functional districts and land use types [35–38] and to improve the accuracy of population spatialization [1,9,15].

Research on population spatialization has led to considerable improvements. In spatial resolution, some researchers have achieved the building-scale population spatialization simulation [9,10]. In temporal resolution, scholars have not only implemented the seasonal population spatialization [39], but also simulated the spatial distributions of daytime and nighttime populations [40], and achieved nearly real-time population estimation using mobile phone signaling data [33] and Twitter data [30]. In the composition of populations, researchers have explored the relationship between different populations (work, home, and mixed populations) and building volume derived from laser radar data [27] and distinguished residents and non-residents using Twitter data [30]. Although it is rare, some experts have tried to spatialize different age populations, such as Alegana et al. [41], who predicted the proportion of the under-five population in 1 km × 1 km in Nigeria. Zhao et al. [42] generated the 100 m × 100 m maps of elderly individuals and children in the area within the Beijing Fifth Ring Road, Beijing, China. The spatialization of different age populations is helpful for decision makers to make more targeted plans, which has broad application prospects, for example, to identify vulnerable people in the COVID-19 pandemic, and to optimize the location of public services and commercial facilities. This is conducive to the further optimization of urban resource allocation and promotes the sustainable development of the city. Current auxiliary variables for population spatialization (e.g., land use, nighttime light, and geographic information data) better indicate the total amount than the age structure of the population in a spatial unit. POI data have the potential to simulate the spatial distribution of different age populations. Previous studies have demonstrated that POI data have different characteristics to attract populations, which means that the number of people gathered near different POIs varies [43–45]. This is because POI data represent all kinds of infrastructures and reflect the supply and demand of humans. There is some relationship between the spatial distributions of POI data and population demand, while the quantity and type of population demand depend on the composition of the populations. Therefore, we can assume that the number and spatial distributions of different attribute populations can be simulated by POI type and quantity on a certain scale. Zhao et al. have used multiple auxiliary variables (containing POI data) in a study of different ages population spatialization [42]. However, previous studies applied POI data to spatialize different age populations subjectively, and the lack of research on the relationship between the two variables, the relationship between spatial distributions of POI data and different age populations, is still unclear; whether it can be...
used as an auxiliary variable for the different age population spatialization remains to be explored.

Therefore, this study puts forward the scientific questions as follows: (1) What is the relationship between different types of POI and different age populations? (2) Can POI data be used as an auxiliary variable for the spatialization of different age populations? To explore the above questions, we collected and sorted out the number of different age populations and POI data in 2846 county-level administrative units of the Chinese mainland in 2010. Then, we divided the research data into different regions, population-based size cities, and economic zones. Finally, we used analysis methods combining conventional statistics (Pearson correlation coefficient) and spatial statistics (Geodetector) to explore the relationship between population and POI. The research results can provide a basis for the spatialization of different age populations. Although using POI data comes with high costs, the POI data is dynamic and available for a high frequency. On the contrary, the population census is established every 10 years and has a significant cost, thus our research will provide a potential way to estimate the dynamics of population structure distribution in large areas. The results of this study will contribute to the development of the dynamics of population structure distribution in large areas.

2. Materials and Methods

2.1. Data and Study Area

Population data were obtained from “Tabulation on the 2010 population census of the People’s Republic of China” [46], which is the number of permanent residents, namely the number of people who have lived locally for more than six months. In this paper, the total population and different age populations are represented by the indicators shown in Table 1. Due to the lack of population data of Taiwan Province, Hong Kong, and Macao, these regions were excluded from the analysis, which covers mainland China only. Finally, we summarized and obtained five population indicators of 2846 county-level units based on county-level administrative units.

Table 1. Population indicators.

| Type   | Code  | Count            | Description              |
|--------|-------|------------------|--------------------------|
| General| POPG0 | 1,345,011,291    | Total population         |
| Age    | POPA1 | 222,896,304      | 0–14 years old population (Child) |
|        | POPA2 | 429,961,497      | 15–34 years old population (Young) |
|        | POPA3 | 513,273,500      | 35–59 years old population (Middle-aged) |
|        | POPA4 | 178,879,990      | 60 years old and older population (Old) |

POI data were obtained from AutoNavi Map (https://www.amap.com/) (accessed on 2 October 2021) and the Chinese web map (Gaode Map in Chinese) in 2015. POI data represent objects in space, such as parks, schools, and hospitals [47]. The total number of POI data is 16,602,978 and the data contains 44 layers. We merged POI data into 8 categories and 28 subcategories (Table 2), and then counted the number of POI within county-level administrative regions. China’s administrative regions are divided into four levels from the top own, containing provincial-level administrative regions, prefecture-level administrative regions, county-level administrative regions, and township-level administrative regions. Excluding the regions with missing data, the number of county-level administrative regions was up to 2846 in 2010, and the demographic data were easy to obtain. To facilitate the comparison of POIs and population data, we use the county-level administrative regions in 2010 for the data summary. China’s population census, which is conducted every ten years, is the most authoritative population survey data. The survey is mainly composed of county-level administrative units. The latest issue is 2020, but the release of the detailed data may take time, so we do not yet have access to the 2020 census data. 2010 is the best data available at present. Similarly, detailed POI data began to be available after 2015, but at present, the data are mainly controlled by some network companies. POI data on a national
scale is not free and the cost is very high. The purpose of this study is to explore and understand the relationship between population and POIs based on the most authoritative population survey data and POI data to provide a scientific basis for further simulating the spatial distribution of population structure.

**Table 2.** The categories of POI.

| Code  | Count     | Category                                                                 |
|-------|-----------|---------------------------------------------------------------------------|
| POIAll| 16,030,423| All POI                                                                   |
| POIF  | 457,612   | Finance                                                                   |
| POIF1 | 80,905    | Insurance companies, financial and insurance institutions                |
| POIF2 | 202,934   | Bank                                                                      |
| POIF3 | 173,771   | Bank-related, ATM, securities company                                     |
| POIT  | 1,118,565 | Traffic                                                                   |
| POIT1 | 518,681   | Transportation facilities (airports, train stations, bridge, traffic place-name, road ancillary facilities, transport facilities services) |
| POIT2 | 536,872   | Vehicle services (motorcycle service, car service, car maintenance)      |
| POIT3 | 63,012    | Automobile sales                                                          |
| POIA  | 1,022,010 | Administration                                                             |
| POIA1 | 889,289   | Administrative management (regional and municipal governments, district and county governments, government agencies, and social organizations) |
| POIA2 | 132,719   | Public security organs                                                     |
| POIE  | 614,441   | Education                                                                 |
| POIE1 | 215,371   | Primary schools and kindergartens                                          |
| POIE2 | 58,769    | High school                                                                |
| POIE3 | 82,375    | Colleges and universities                                                  |
| POIE4 | 257,923   | Scientific and cultural services                                           |
| POIH  | 686,630   | Health                                                                     |
| POIH1 | 136,979   | Hospital (general hospitals, specialized hospitals)                        |
| POIH2 | 549,650   | Health care services                                                       |
| POIL  | 6,703,828 | Life                                                                       |
| POIL1 | 495,144   | Residential communities                                                    |
| POIL2 | 5,808,468 | Shopping services, catering services, life service                         |
| POIL3 | 400,214   | Sports leisure services, scenic spots                                      |
| POIO  | 1,564,313 | Office                                                                     |
| POIO1 | 1,445,752 | Company enterprise                                                         |
| POIO2 | 60,843    | Business residence                                                        |
| POIO3 | 57,717    | Office building                                                            |

To explore the differences between differently sized cities, we divided 294 prefecture-level cities into four population-based sizes (Figure 1), according to the 2010 sixth national population census data of China, and we refer to the city classification of Dou et al. [48], as follows: megacities (more than 5 million population), large cities (1 to 5 million population), medium cities (0.5 to 1 million population), and small cities (less than 0.5 million population).

To compare regional differences, cities within China were divided into four economic zones, according to China’s Economic Geographical Zoning Scheme [49] (Figure 1): (1) the northeastern zone of China (NEC) containing Heilongjiang, Jilin, and Liaoning; (2) the eastern zone of China (EC) containing Beijing, Shanghai, Tianjin, Shandong, Guangdong, Jiangsu, Hebei, Zhejiang, Hainan, and Fujian; (3) the central zone of China (CC) containing Shanxi, Henan, Anhui, Jiangxi, Hubei, and Hunan; and (4) the western zone of China (WC) containing Chongqing, Sichuan, Yunnan, Guizhou, Gansu, Shaanxi, Qinghai, Ningxia, Guangxi, Xinjiang, Tibet, and Inner Mongolia. In addition, we divided the Chinese mainland into the southeastern (SEC) and northwestern (NWC) regions, indicated by the Hu Huanyong Line. The Hu Huanyong Line (Heihe-Tengchong line), proposed by Chinese geographer Hu Huanyong (1901–1998) in 1935, is an important division line of population density in China which has been recognized by demographers and geographers throughout the world [50,51]. The natural environment and urbanization levels of the SEC and NWC regions indicated by the Hu Huanyong Line are quite different. The SEC region indicated
by the Hu Huanyong Line is dominated by plains, hills, water network, karst, and Danxia landforms, accounting for 43.8% of the country but 95% of the population. The population density in NWC is very low, and the area is mostly grasslands, deserts, and snow areas.

Figure 1. Locations of provinces and cities in China.

2.2. Methods

2.2.1. Pearson Correlation Coefficient

Correlation analysis is an analytical process to explore whether there is a certain association between two continuous variables, and to explore direction and degree of association. The Pearson correlation coefficient is a common measure of correlation which evaluates the linear relationship between two variables, $X$ and $Y$, and the range of the value ($-1, +1$). The greater the absolute value of the correlation coefficient, the stronger the correlation [52].

2.2.2. Geodetector Analysis

The Geodetector is a statistical tool that can evaluate spatially stratified heterogeneity and identify its determinants [53]. The Geodetector includes a risk detector, factor detector, ecological detector, and interaction detector [54]. This study applied the factor detector and interaction detector to evaluate the relationship between POI and different population age groups. The factor detector can detect the explanatory power of each independent variable. For the $q$ value in Geodetector, which quantifies explanatory power, the formula is as follows:

$$q = 1 - \frac{\sum_{i=1}^{L} N_i \sigma^2_i}{N \sigma^2}$$  \hspace{1cm} (1)

where $\sigma^2$ and $N$ denote the variance and number of units of population indicators of the study area, the study area is stratified into the $L$ stratum $i = 1, 2, \ldots, L$ denotes the number of stratum, and $N_i$ and $\sigma^2_i$ denote the number of units of stratum $i$ and variance of population indicators of the stratum $i$, $q \in [0, 1]$, $q = 1$ means the independent variable completely controls the population indicators, and $q = 0$ means the independent variable is completely unrelated to the population indicators.
The interaction detector was applied to explore whether two driving variables, A and B, when taken together, weaken or enhance one another, or whether they are independent in developing the spatial pattern of the dependent variable. Table 3 shows the criterion of the interaction relationship.

### Table 3. The interaction relationship between two variables.

| Criterion | Interaction          |
|-----------|----------------------|
| q (A∩B) > q (A) and q (B) | Enhance, bivariate   |
| q (A∩B) > q (A) + q (B) | Enhance, nonlinear   |
| q (A∩B) < q (A) + q (B) | Weaken               |
| q (A∩B) < q (A) or q (B) | Weaken, univariate   |
| q (A∩B) < q (A) and q (B) | Weaken, nonlinear    |
| q (A∩B) = q (A) + q (B) | Independent          |

When we used Geodetector, we divided the study area into several zones. To determine the most appropriate quantity of the zones and division method of the study area, we performed many tests and calculations by the "GD" package in the R, and division methods containing “Natural breaks”, “Equal interval”, “Quantile”, “Geometric” and “SD” [55]. The so-called most appropriate quantity and division method mean that independent variables have the highest explanatory power for dependent variables when the study area is divided into certain quantity zones and a certain division method is selected.

### 3. Results

#### 3.1. Descriptive Statistics of the Spatial Distribution of Different Age Populations in China

There was obvious spatial heterogeneity in the population distribution in China in 2010 (Figure 2). To show the spatial patterns more clearly, the value of population was divided into six categories by the natural breaks method. The spatial pattern of POPG0 shows that the total populations in China were higher in the southeastern region, indicated by Hu Huanyong Line, than in the northwestern. Table 4 displays the descriptive statistics of population indicators. The lowest nonzero value of POPG0 appeared in Zanda County, Ngari Prefecture, Tibet, with the number of only 6883 people. The highest POPG0 value was in Dongguan City (8,220,207 people). Dongguan is the only prefecture-level city without county-level regions, with a developed economy and large population, so the POPG0 value was the largest in statistics when we compared it with other county-level regions. From the perspective of county-level administrative regions, the Pudong New Area of Shanghai was the most populous county-level administrative unit, with up to 5,044,430 people, about 733 times that of Zanda County (Figure 2).

### Table 4. Descriptive statistics of the population indicators.

| Indicator Name | Observation | Mean S.D. | Min | Median | Max |
|----------------|-------------|-----------|-----|--------|-----|
| POPG0          | 2846        | 466,062   | 397,625.8 | 0     | 380,106 | 8,220,207 |
| POPA1          | 2846        | 77,341    | 64,692.96 | 0     | 62,388  | 804,756   |
| POPA2          | 2846        | 149,189   | 169,257.2 | 0     | 113,880 | 4,756,885 |
| POPA3          | 2846        | 178,096   | 143,897.5 | 0     | 147,425 | 2,492,729 |
| POPA4          | 2846        | 62,068    | 50,592.06 | 0     | 49,679  | 705,947   |

The largest value of each age population indicator generally appeared in the southeastern coastal areas, while the lowest nonzero value appeared in underdeveloped areas, such as the northeastern and northwestern regions (Figure 3). We also divided different age populations into six categories by the natural breaks method. The minimum nonzero values of POPA1, POPA2, POPA3, and POPA4 appeared, respectively, in Shangganling District, Yichun City, Heilongjiang Province (1420 people); Urho District of Karamay City, Xinjiang (2262 people); Zanda County of Ali Region, Tibet (1907); and Zanda county (538 people).
The maximum values of POPA1, POPA2, POPA3, and POPA4 appeared in Dongguan (680,643), the Bao’an District of Shenzhen City (2,903,104), and the Pudong New Area of Shanghai City (1,951,754, 705,947).

**Figure 2.** Spatial pattern of the total population in the county-level administrative regions on the Chinese mainland.

**Figure 3.** Spatial patterns of the different age populations in the county-level administrative regions on the Chinese mainland. POPA1, POPA2, POPA3, and POPA4 represent populations aged 0–14 years old, 15–34 years old, 35–59 years old, and 60 years old and older, respectively.
Young (POPA2) and middle-aged populations (POPA3) constituted the majority of China’s population (Figure 4). POPA3 and POPA2 accounted for 38% and 32%, respectively, while the proportion of POPA1 (child) was 17%, and the old population accounted for only 13%.

**Figure 4.** The proportion of the different age populations.

Due to space limitations, we only show the spatial pattern of the POIall amount classified by the natural breaks method (Figure 5). The number of POI indicators in each county-level administrative region is counted. Table 5 shows the descriptive statistics of all kinds of POI indicators summarized by county administrative regions. Except for the number of POIall in some county-level administrative areas that are zero, the minimum nonzero value of POIall appeared in the Xingshan District, Hegang City, Heilongjiang Province, with only 108 POI points, while the Pudong New Area in Shanghai was the county-level administrative region with the most POI, with 90,280 POI points, 835 times that of the Xingshan district. The spatial distributions of POIall and populations were similar. Generally, the number of POIall and the populations in the southeastern region were higher than in the northwestern region, and there was an overlap between the high-value area of population and POIall. POIL (POIs which are related to daily life) accounted for the highest proportion of POIall (42%), while the POIs related to finance, health, and education are less, accounting for only 3–4% (Figure 6).

### 3.2. Pearson Correlation Coefficient

For different regions, POIE2 (high school) was an important explanatory variable of population indicators (Figure 7). Figures 7–9 show the scatter plots between POI and population. All plots and Pearson correlation coefficients show positive, moderate-to-strong linear correlations. However, there may also be substantial scatter around the regression line and the linear relationship may not be appropriate in all parts of the data space. The POI with a strong correlation with population indicators was generally POIE2, and only the most important POI indicators of POPA2 were POIall and POIH. Except for the relatively weak correlation with POPA1, the correlation between POIE2 and POPG0, POPA3, and POPA4 were generally greater than 0.8, while the R-squared was more than 0.6, and some reached more than 0.75, which showed that a single POIE2 variable can well explain the spatial distribution of population indicators.
Figure 5. The spatial pattern of POIall in the county-level administrative regions on the Chinese mainland in 2015.

Figure 6. The proportion of different POIs. (a–h) represent POIall, POIF, POIT, POIL, POIO, POIA, POIE, and POIH, respectively.
Table 5. Descriptive statistics of the POI indicators summarized by the county-level administrative regions.

| Indicator Name | No. of Counties | Mean No. POIs of All Counties | S.D. | Min | Median | Max |
|----------------|-----------------|-------------------------------|------|-----|--------|-----|
| POIall         | 2846            | 5562                          | 6653.18 | 0  | 3836   | 168,085 |
| POIF           | 2846            | 28.07                         | 74.89 | 0  | 52     | 1535 |
| POIF1          | 2846            | 70.41                         | 112.84 | 0  | 26     | 2383 |
| POIF2          | 2846            | 60.3                          | 112.84 | 0  | 26     | 2383 |
| POIF3          | 2846            | 388.1                         | 665.08 | 0  | 206    | 13,832 |
| POIT           | 2846            | 180                           | 441.11 | 0  | 69     | 8756 |
| POIT1          | 2846            | 186.3                         | 240.76 | 0  | 117    | 5700 |
| POIT2          | 2846            | 21.86                         | 34.31 | 0  | 9      | 457 |
| POIT3          | 2846            | 354.6                         | 321.61 | 0  | 272    | 6505 |
| POIA           | 2846            | 308.6                         | 277.33 | 0  | 237    | 5452 |
| POIA1          | 2846            | 46.05                         | 48.10 | 0  | 34     | 1053 |
| POIA2          | 2846            | 213.2                         | 292.79 | 0  | 133    | 5275 |
| POIE           | 2846            | 74.73                         | 74.01 | 0  | 57     | 1112 |
| POIE1          | 2846            | 20.39                         | 16.99 | 0  | 16     | 249 |
| PO IE2         | 2846            | 28.58                         | 58.66 | 0  | 10     | 1337 |
| POIE3          | 2846            | 89.49                         | 177.06 | 0  | 41     | 3531 |
| POIH           | 2846            | 238.2                         | 290.80 | 0  | 155    | 8545 |
| POIH1          | 2846            | 47.53                         | 48.07 | 0  | 33     | 799 |
| POIH2          | 2846            | 190.7                         | 251.97 | 0  | 119    | 7746 |
| POIL           | 2846            | 2326.1                        | 3661.81 | 0  | 1184.5 | 95,130 |
| POIL1          | 2846            | 171.8                         | 376.76 | 0  | 49     | 6296 |
| POIL2          | 2846            | 2015.4                        | 3160.66 | 0  | 1059.5 | 88,530 |
| POIL3          | 2846            | 138.9                         | 199.10 | 0  | 72     | 3328 |
| POIO           | 2846            | 542.8                         | 1323.04 | 0  | 207    | 22,994 |
| POIO1          | 2846            | 501.65                        | 1211.88 | 0  | 194    | 31,514 |
| POIO2          | 2846            | 21.11                         | 82.74 | 0  | 3      | 1624 |
| POIO3          | 2846            | 20.03                         | 56.23 | 0  | 4      | 1235 |

The correlation of POI with different population indicators varied. The correlation between POPA1 and POI was 0.675–0.714, while the correlations between other population indicators and POI were mostly more than 0.8.

In differently sized cities, POI was always the variable with the largest correlation of population indicators, and POIE2 was the most important explanatory variable (Figure 8). The results of the correlation analysis showed that the correlation between POI and different population indicators was higher than 0.7, and some reached more than 0.9. Among the POI with the largest correlation to population, POIE2 appeared 17 times, accounting for 85%.

The correlation of POI with population indicators was generally proportional to the size of the city. Except for POPA4, the results of other population indicators and POI showed that the correlation coefficient was the highest in megacities, and gradually decreased with the reduction of city size.

The correlation between POI and population indicators varied from zone to zone (Figure 9). The range of correlation coefficient in the CC zone was 0.694–0.834, which was generally lower than other zones, while it was generally highest in the NEC zone, with a range of 0.826–0.908.

POIE2 was an important variable that was highly correlated with population indicators in different economic zones. Among the POI indicators with the largest correlation, POIE2 accounted for 90%.
Figure 7. Scatter plots of POI and population in different regions. In each subgraph, the red line represents the univariate linear regression line, and the POI in the x-axis is the POI with the highest correlation to the population indicators. All correlation coefficients were significantly different from zero ($p < 0.01$).

Figure 8. Scatter plots of POI and population in differently sized cities. In each subgraph, the red line represents the univariate linear regression line, and the POI in the x-axis is the POI with the highest correlation to the population indicators.

The points that represent single administrative units that are very far from the regression line in Figures 7–9 and deserve our attention. In the “NWC region” in Figure 7, several outliers appear at the top left of the scatter plots of POPA1 and POIE2, the outliers are from
Yunnan, Guangxi, and Guizhou in China. In the scatter plot of POPA1 and POIE2 of “Large city” in Figure 8, there are two obvious outliers, which are from Shantou City, Guangdong Province. A series of outliers also appear in the scatter plot of POPA1 and POIE2 in the “Small city” of Figure 8. Most of these points are located in Yunnan, Guangxi, and Guizhou. These points coincide with the outliers of POPA1 and POIE2 in the NWC region of Figure 7. The outliers of the scatter plots of POPA1 and POIE2 in the WC region in Figure 9 are also from Yunnan, Guangxi, and Guizhou. These outliers show that there is a large number of children, but the ratio of middle schools to children is lower than that in other regions.

3.3. Geodetector Analysis
3.3.1. Factor Detector and Interaction Detector for Different Regions

For the different regions, the dominant variable and interaction variables are shown in Figure 10. It is worth mentioning that the region in this paper specifically refers to the whole country and the southeastern and northwest regions indicated by the Hu Huanyong Line. The economic zones refer to four economic zones divided by the Chinese government according to different social and economic development levels. Cities are the prefecture-level administrative units in China and are an integral part of regions and economic zones.
Figure 10. Factor detector (a) and interaction detector (b) of POI to population indicators in different regions. Max_Q represents the q value of the POI indicator with the strongest explanatory power for the population indicator in the Geodetector. All of the results were significantly different from zero \((p < 0.01)\). ‘All’ denotes all regions.

POIE2 was the strongest explanatory variable in the SEC, which was consistent with the whole country. The strongest explanatory variables in the NWC were mostly health POI, and the strongest explanatory variables of POPG0, POPA2, and POPA3 were POIH (Health), POIH2 (health care services), and POIH (Health) respectively. For the interaction of two variables, the strongest explanatory interaction in the SEC was POIall∩POIE2, and the most common variables in the strongest explanatory interaction combination of NWC were health POIs. The strongest interactions of POPG0, POPA1, POPA2, and POPA3 were POIE2∩POIH2, POIH2∩POIO3, POIT1∩POIH, and POIE2∩POIH2, respectively.

Generally, POI data had good explanatory power for different age populations. For POPA1, POPA2, and POPA3, Max_Q values of dominant explanatory variables were higher than 0.7, and Max_Q values of interactions were generally higher than 0.8. This showed that POI can be used as an effective auxiliary variable for different age populations’ spatialization. In addition, we also found that the Max_Q value of dominant explanatory variables and interaction variables of POI on POPA1 was generally lower than that on other age populations, while the explanatory power of POI on POPA3 was generally higher than that of other age populations, which suggested that POI had a low ability to explain the spatial distributions of the child-age population and a high ability to explain the spatial distributions of the middle-aged population.

3.3.2. Factor Detector and Interaction Detector for Differently Sized Cities

The explanatory power of POI to the population was directly proportional to the size of the city (Figure 11). For the dominant explanatory variable, the maximum value of Max_Q (0.877) appeared in megacities, while the minimum value (0.483) appeared in small cities. The means of Max_Q value in megacities, large cities, medium-sized cities, and small cities were 0.773, 0.649, 0.643, and 0.615, respectively. This indicated that with the decline of city size, the explanatory power of POI to the population gradually decreased. For the interaction, the maximum value of Max_Q (0.930) also appeared in megacities, and the minimum value (0.577) appeared in small cities. The means of Max_Q
in megacities, large cities, medium-sized cities, and small cities were 0.878, 0.760, 0.723, and 0.680, respectively. This not only showed that the explanatory power of the two-variable interaction on population was directly proportional to the city size, but also indicated that the two-variable interaction can better explain the spatial distributions of the population.

Figure 11. Factor detector (a) and interaction detector (b) of POI to population indicators in differently sized cities.

The explanatory power of a single POI or a combination of POIs to POPA1 was lower than that of other age groups. For the Max_Q value of the dominant explanatory variable, the mean Max_Q value of POPA3 was the highest (0.729), while the mean Max_Q value of POPA1 was the lowest (0.577). The Max_Q value of interaction also showed that the mean Max_Q value of POPA1 was the lowest (0.69).

POIE2 was an important explanatory variable for the population. In the dominant explanatory variable, 85% of the maximum explanatory variable was POIE2, and in the interaction, 85% of the maximum interaction included POIE2.

3.3.3. Factor Detector and Interaction Detector for Population Indicators in Different Economic Zones

The explanatory power of POI to population was the highest in NEC and the lowest in WC (Figure 12). For the dominant explanatory variable, although the maximum value of Max_Q (0.812) appeared in EC, the mean of Max_Q value in NEC was 0.717, while WC was the smallest, at only 0.620. For the mean value of the Max_Q of interaction, NEC (0.757) > EC (0.765) > CC (0.709) > WC (0.698).

POIE2 was always an important explanatory variable in every region. In the dominant explanatory variable, all of the maximum explanatory variables were POIE2, and in the interaction, 95% of the maximum interaction included POIE2.
4. Discussion

4.1. Complex Relationships between the POI and Different Age Populations Exist at Various Spatial Scales

The results show that there are positive, moderate-to-strong linear correlations between POI and population indicators, but the linear relationship may not be appropriate in all parts of the data space. Recent studies often use nonlinear models for POI-based population simulation [1,15,56]. In addition, we found that the explanatory power of POI combinations on population indicators is generally higher than that of a single variable. Therefore, when carrying out population spatialization, in addition to considering the application of nonlinear models, we can try to combine different POIs as predictors to get better simulation results.

The explanatory power of POI to different age populations is positively correlated with the level of urban economic development. For the differently sized cities, the scale is directly proportional to the explanatory power of POI to populations. For the different regions, the explanatory power of POI in the economically developed eastern regions is higher than that in the central and western regions. This may be due to the relatively complete facilities in economically developed cities and the high level of population urbanization, as the more populations are concentrated in urban areas, the more likely people are to access various facilities, and thus the spatial distributions of the population are more easily simulated by POI points. While there are fewer POIs in smaller cities, the distribution of POIs is sparse and it is not easy for people to access various facilities, so the spatial distributions of the population are difficult to be captured by POIs. Therefore, population spatialization using POI in underdeveloped areas may not yield ideal results.

POI has a lower explanatory power for the child and old population, while it has higher explanatory power for the middle-aged and young population. This indicates that the spatial distributions of POIs mainly overlap the activities and needs of the middle-aged and young population. This may be due to the strong activity of the young and middle-aged population and the greater demand for various services.

The relationship between POI and population may be affected by local childbearing customs. We found that there is a large number of children in Shantou city in eastern China and in the Guangxi, Guizhou, and Yunnan provinces in the western part of China, but the ratio of POI to the number of children in middle school is lower than in other locations. From 1982, “Family planning” became a basic state policy for China, which is to control...
the population in a planned way. The main content of “Family planning” is to advocate late marriage, late childbearing, less birth, and better birth. The Chinese government once implemented the “Family planning” policy to reduce the birth rate, which has made many Chinese families have only one child. However, there may be local deviations from this policy. These regions have different ideas of childbearing and a strong willingness to have more children [57].

POI can be considered as an explanatory variable or as a response explanatory variable. It may be that POI attracts specific people, resulting in population spatial distribution, or it may be that specific POI is planned according to the spatial distribution of different populations. In terms of the correlation analysis, we only look at relationships between variables. The distinction between response and explanatory variables is not relevant in the same way as it is for regression or Geodetector analysis. In this study, we hope to take the Geodetector analysis we used for POI as an explanatory variable of the population to explore whether POI can be used to simulate and explain the spatial distribution of the different age populations. We emphasize that this is a statistical explanation that could support predictive analysis. It does not necessarily mean that there is a causal relationship between POIs and population.

4.2. POI Can Be Applied to Predict the Spatial Distributions of Different Age Populations

In the regression analysis and Geodetector results, POI had different explanatory power for different age populations, and explanatory power was generally good. Some scholars have pointed out that different POIs have different characteristics that are appealing to the general population, which means that the number of people gathered near the different POIs varies [43–45]. Our results show that the attractiveness of different POIs to different age groups varies. This may be because different age populations have different needs, they have a certain tendency in choosing POI, or there are differences in the frequency of choosing a certain type of POI. Therefore, POI can be used to reflect the spatial distributions of different age populations. POIE2 is the most common variable with the strongest explanatory power. This is similar to the results of Yang et al. [1], where the POI of education facilities had the largest explanatory power to the population. The highest explanatory power variable is POIE2, rather than POIs, which suggests that when using POI data for population spatialization, the best prediction effect cannot be obtained by using all POI, and the researcher can further subdivide POI and simulated populations by more relevant categories. In addition, POIE2 generally has good explanatory power for different age populations. This shows that an ideal population spatialization result can be obtained by using POIE2 in certain situations. It is worth noting that the number of POIE2 is not large (58,769), accounting for only 0.354%. This indicated that a relatively accurate population spatialization result can be simulated with a small number of POIE2 points.

4.3. Limitations of the Study

Our study also has some limitations. Firstly, due to the lack of finer-scale data, we carried out research at the county level. In the future, finer scale research at the functional scale can be undertaken in order to better distinguish the relationship between different populations and POI. Secondly, the categorization of POIs (44 categories) was already determined by the data provider. We have combined these into 28 related categories. The categorization and aggregation of the POIs could be an issue that may present challenges to researchers. Different data providers may provide different categorization systems. Nevertheless, our results will still be informative to other researchers. Thirdly, due to the availability of data, we cannot make a multi-period comparison. Data from multiple time periods could lead to further insights. However, detailed POI data on a national scale is costly and the availability and quality of multi-period datasets can be problematic. Considering the rapid development of the social economy and network technology, as well as the enhancement of awareness of data sharing, it may become easier to obtain detailed POI data in the future. The results of this paper can also be used as a reference for similar future studies to com-
pare the changes in the relationship between POI and population. Finally, we assume that
the relationships between POIs in similar categories and populations are consistent [56],
which ignores the different quantitative and qualitative characteristics of POI (e.g., scale,
accessibility, etc.). Meanwhile, some of the POIs require many years of planning and
implementation (e.g., airports, universities), and their potential is often not fully realized at
the time of launch. These differ from the POIs that can provide services quickly (e.g., ATMs,
catering services), and the attraction of POIs whose potential is not fully realized may
lead to a time-lagged relationship with population. In further research, we should take
more account of the quantitative and qualitative characteristics of POIs, the changes in the
relationship between POI and population over time, and more refined POI classifications
to obtain more detailed results.

5. Conclusions

POI data with rich information have great potential to simulate the spatial distribution
of different age populations, but the relationship between spatial distributions of POI and
different age populations is still unclear, and whether it can be used as an auxiliary variable
for the spatialization of different age populations remains to be explored. Therefore,
this study collected and sorted out the number of different age populations and POIs
in 2846 county-level administrative units of the Chinese mainland in 2010, divided the
research data by region and city scale, and explored the relationship between the different
age populations and POIs.

We found there is a complex relationship between POI and different age populations.
Firstly, there are positive, moderate-to-strong linear correlations between POI and pop-
ulation indicators. Secondly, the explanatory power of POI to different age populations
is positively correlated with the urban economic development level. Finally, POI has a
lower explanatory power for the child and old populations, while it has higher explanatory
power for the young and middle-aged populations. In addition, the results showed that
POI can be applied to predict the spatial distributions of different age populations. There
was a significant and strong correlation between POI and different age populations, and the
relationships varied. POI generally had good explanatory power for different age popu-
lations. A small number of a certain kind of POI can be used to effectively simulate the
spatial distributions of different age populations, which can greatly improve the efficiency
of obtaining spatialization data of different age populations and greatly save on costs.

Our study explores the relationship between the spatial distributions of POI and
different age populations, and it demonstrates that POI can be used as an auxiliary variable
for the different age population spatialization. The results can provide data support for the
spatialization of different age populations in the future and ideas for the spatialization of
the population with different attributes.

Author Contributions: Methodology, Yiyi Huang and Guoqin Zhang; writing—reviewing and
editing, Tao Lin, Wei Zhu, Nicholas A. S. Hamm; software, Yiyi Huang, Guoqin Zhang, and Yuqin
Liu; conceptualization, Tao Lin and Yiyi Huang; visualization, Junmao Zhang and Xia Yao. All authors
have read and agreed to the published version of the manuscript.

Funding: This research was funded by the Major Special Project—the National Natural Science
Foundation of China (No. 41871167), the China High-Resolution Earth Observation System (30-Y30F06-9003-20/22),
and the Bureau of International Cooperation, the Chinese Academy of Sciences (GJHZ202118).

Institutional Review Board Statement: Not applicable.

Informed Consent Statement: Not applicable.

Data Availability Statement: Not applicable.

Conflicts of Interest: The authors declare no conflict of interest.
