Exploration of the Relationship between Travel Activity Chains and Self-Assessed Health of Rural Community Residents: Based on Multi-Source Data and Health Survey Questionnaire

Kehao Li, Xiu Yang, Hu Yumin, Chao Liu
Affiliation: Wuhan University, Wuhan, China.
Email: lkhwork04@163.com, yangxiu2022@whu.edu.cn

Abstract
At present, most of China's population still lives in rural areas, and the allocation of health facilities in rural areas is generally weaker than that in urban areas. The factors affecting the healthy life of rural residents need to be studied. This study took six rural communities in Pujiang County, Chengdu, Sichuan Province as the research object, and randomly selected 172 residents as samples to investigate the relationship between the characteristics of daily travel activities and self-rated health level of rural community residents, in order to provide a scientific basis for the construction of healthy rural communities and their surrounding supporting facilities. In view of the difficulty of data collection in rural areas, such studies are relatively rare in China. The research comprehensively uses multi-source data, including GPS data obtained by wearable positioning equipment, health questionnaire data, POI data of Amap and building vector data of MapWorld. Firstly, this study uses AT-DBSCAN algorithm to cluster the travel trajectory points to obtain the location information of residents' stay points between 8:00-22:00, and divides the study area into 120m long grids. The TF-IDF algorithm is used to obtain the functional attributes of the grid based on the classified POI data, and then infer the type of stay point activities. For rural areas without POI data, the daily travel activity chain of rural community residents is constructed by combining building vector data. Then, the Word2Vec word embedding model is used to vectorize the activity chain data, and the K-means clustering analysis is used to obtain different daily travel activity patterns. Finally, combined with the residents' self-rated health data obtained from the health questionnaire, the Nonparametric Tests was used to analyze the differences between the self-rated health levels of different travel activity modes. The results show that there are significant differences in self-rated health level among rural community residents with different travel activity modes, which provides a reference for rural community planning and health management.

Keywords: Travel activity, characteristics, Rural community, Self-rated health.

Introduction

With the development of social economy, people pay more and more attention to the healthy lifestyle. In addition to basic physical and mental health, residents' daily travel contact with the living environment is also closely related to healthy life. Studying the relationship between the environmental characteristics and health status of residents' travel activities will help managers create a healthy living environment from the planning and design level. At present, the research on the relationship between the characteristics of travel activities and residents' health is relatively rich in urban areas, but insufficient attention is paid to rural areas. The prediction released by the United Nations in 2018 shows that China's urbanization rate will be 70.6% and 80.0% in 2030 and 2050, and the population in rural areas will be about 399million and 272million(United Nations Department of Economic and Social Affairs, Population Division, 2019)ΔAt the same time, China's 2021 census shows that 36.11% of the population still lives in the built-up environment in rural areas (National Bureau of Statistics of China, 2021)Δand the allocation of health facilities in these areas is generally weaker than that in urban areas, which significantly affects the health level of rural residents. In the past, China's research on rural health mostly focused on medical services (Gao *et al.*, 2024; Yang *et al.*, 2024) and environmental factors, but the link between residents' daily travel activities and health was rarely involved. In fact, daily travel activities have a direct or indirect impact on rural residents' access to medical care, leisure and social services, which can't be ignored in rural health research.

In view of this, this study selected some residents from six rural communities in Pujiang County, Chengdu, Sichuan, China to investigate and analyze the relationship between their daily travel activity characteristics and their self-rated health level, focusing on the characteristics of stay activity area and activity time allocation, so as to provide a scientific basis for the construction of healthy rural communities and their surrounding supporting facilities. Considering the difficulty of data collection in rural areas, relevant research in this field is relatively rare. In this study, multi-source data, including GPS data obtained by wearable positioning equipment, health questionnaire data, POI data of Amap and building vector data of MapWorld, were comprehensively used to explore the potential relationship between travel activity characteristics and health level, and provide beneficial reference for improving the health status of rural residents, optimizing rural community planning and health management.

Literature Review

Relationship between travel activity characteristics and health

Travel activity characteristics mainly include travel time characteristics, travel space characteristics, activity type, travel frequency, travel mode, travel destination, etc. Many studies have confirmed the correlation between travel activity characteristics and health, mainly through the built environment as the medium, from the aspects of physical activity promotion, environmental exposure, nature of activity sites, environmental healing (Frank *et al.*, 2019; Mahmoudi, 2022; Sharma and Jain, 2023). Choosing an active travel mode that can promote physical activity (such as walking, cycling, etc.) and increasing the travel time allocation of physical activity have many impacts on health promotion (Ali *et al.*, 2020, 2024; Dėdelė *et al.*, 2020; Foley *et al.*, 2018). Environmental exposure in travel activities has significant effects on health(Frank *et al.*, 2019; Poom *et al.*, 2021) , but such effects have duality. On the one hand, active travel modes such as walking and cycling make individuals more vulnerable to traffic related air pollution and noise, increasing the risk of respiratory and cardiovascular diseases. On the other hand, if the travel path passes through green space, it can have a positive effect on mental health by contacting nature and relieving stress (Le and Poom, 2024) . In addition, the role of environmental exposure and differences in travel mode, time period selection and personal perception vary, and there is a significant socio-economic imbalance (Briggs *et al.*, 2008; Ma *et al.*, 2021) . There are differences in the impact of staying in different places of activity on health. For example, leisure activities in places such as commercialservices, leisure and entertainment (Fancourt D *et al.*, 2021) and natural areas (Aerts *et al.*, 2018; Bratman *et al.*, 2019; Ewert and Chang, 2018) can promote health, while long-term activities in places such as office areas (Rasheed *et al.*, 2021) and industrial areas (Xu *et al.*, 2020) are generally considered to have adverse effects on health. In addition, Zhu and Fan et al. Found that travel mode, travel duration and distance, and destination choice were associated with emotional health (Zhu and Fan, 2018) . Bicycle travel has a psychological healing effect, while public transport travel has the opposite effect. The longer the travel duration, the lower the happiness and the higher the pressure. The positive mood of travel for the purpose of free activities such as leisure and entertainment is higher. The urban blue-green space and healing built environment that travel through have a certain healing effect on people's mental health and can relieve people's pressure and negative emotions (Vert *et al.*, 2020; Zhang *et al.*, 2019). To sum up, the characteristics of travel activities can affect health through a variety of mechanisms, and there is a theoretical basis for mining the correlation between rural residents' daily travel characteristics and health from the perspective of time distribution of travel activities and stay activity areas.

Travel activity characteristics and limitations of health research in rural areas

In the research on the relationship between residents' travel activity characteristics and health, the current literature mainly focuses on the relevant research in urban environment, while the research in rural areas is relatively less. The number and quality of transportation facilities, service facilities and travel activity places in rural areas are generally worse than the average level in cities, and there are also significant differences in the characteristics of residents' travel activities. At present, the research on the characteristics of travel activities in rural areas mostly focuses on travel distance, travel time, travel mode selection, etc. for example, Akinlotan et al. Revealed the differences in distance and time of rural residents in medical or dental care travel by using the data of the 2017 national family Travel Survey (NHTS) , and pointed out that the travel distance of rural residents is more than twice that of urban residents, and the travel time is also significantly longer (Akinlotan *et al.*, 2021) . In addition, Hearn et al. Also showed that the travel time of rural cancer patients is longer, which may be related to the lack of transportation facilities in rural areas (Hearn *et al.*, 2024).

In terms of the relationship between travel activities and health, some studies began to pay attention to the impact of travel activities on health of residents in rural areas. For example, Berga and Ihlström found that rural residents' daily activities and travel behavior were significantly affected by traffic conditions through qualitative interviews with rural residents in Sweden. In particular, the lack of Public Transport Services Limited residents' independent travel ability, thereby affecting their opportunities to participate in social activities and obtain health services (Berg and Ihlström, 2019)∆In addition, McCrorie's research in Scotland found that rural children have shorter sedentary time and longer time of mild physical activity than urban children, which may be related to the more spacious living environment and more opportunities for outdoor activities in rural areas (McCrorie *et al.*, 2020)∆These studies show that there is a complex correlation between the characteristics of travel activities and residents' health in rural areas, but at present, the relevant research is still relatively limited and needs to be further discussed.

In general, the existing literature has a certain foundation in the correlation between the construction of transportation facilities, service facilities, travel activity places and residents' travel activity characteristics in rural areas, but there is still a large gap in the research on the relationship between travel activity characteristics and residents' health. Future research should strengthen the research on the relationship between travel activities and health in rural areas, so as to better guide the traffic planning and health policy-making in rural areas and improve the quality of life of rural residents.

Data preparation

The data used in this study are from a travel trajectory and self-rated health survey conducted by the author's team in Pujiang County, Chengdu, Sichuan Province from July to August 2023 (Fig 1). This study surveyed 172 residents from six rural communities in the region, namely Hongfu Community, Dongyue Community, Shuangsh Community, Bajiaojing Village, Fuxing Community and Mingyue Village, and was conducted in two batches. The number of residents surveyed in each village is shown in Table 1.

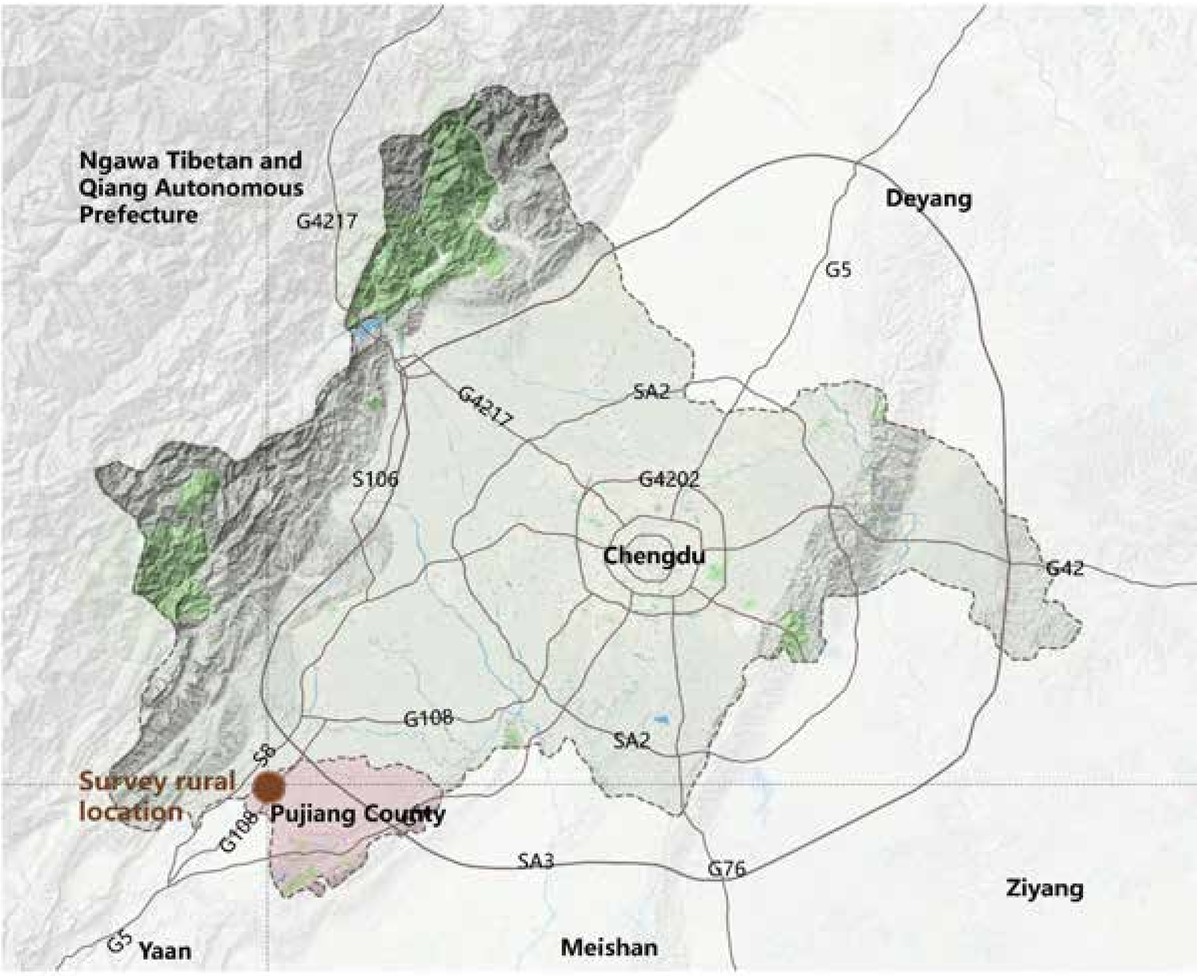

Fig 1. Study area scope and sample source

Table 1. Number of residents surveyed in each village

Number of residents surveyed in each village

Village name	Number of residents surveyed
Hongfu Community	31
Shuangshi Community	33
Dongyue Community	33
Mingyue Village	39
Fuxing Community	32
Bajiao Village	4

GPS positioning data

The GPS data used in this study was obtained by a smart watch with positioning function. The track data recorded by the positioning watch uses WGS 1984 geographical coordinate system, and the positioning accuracy can reach within 30 meters. During the collection process, all positioning watches are set to record trajectory data points every minute, and their information mainly includes longitude and latitude coordinate values, sample numbers, batch number, dates and times (Table 2). During the data collection process, participants were required to wear a positioning watch between 8:00-22:00 within 7 days, and were not allowed to remove the device without special reasons, so as to ensure the continuity of research data collection.

Table 2. GPS data sample

GPS data sample

date	time	sample number	batch number	latitude	longitude
2023-07-12	11:51:00	9617605926	1	30.278349	103.3772225
2023-07-12	11:51:00	9618007196	1	30.28061	103.3820817
2023-07-19	12:13:00	9618007145	2	30.26369	103.3239
2023-07-22	12:12:00	9618007135	2	30.27283	103.3311

As it is difficult for residents to fully cover the entire required number of days for wearing equipment, after screening, a total of 1116 daily behavior tracks were collected, including 584 in the first batch and 532 in the second batch. Fig 2 shows the main distribution range of the trajectories. Most of the data are concentrated in Pujiang County, Qionglai city and Mingshan District, Ya'an City.

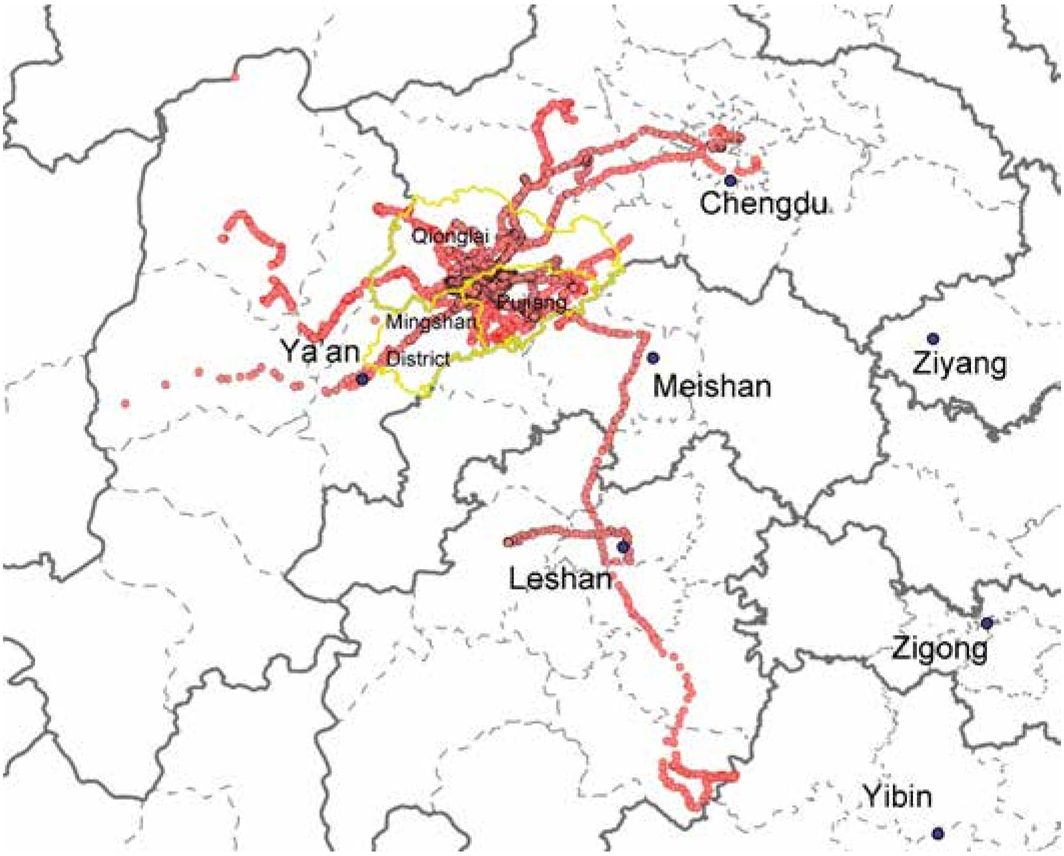

Fig 2. overall distribution of travel trajectories data

Health questionnaire survey data

Self-rated health has been proved to be able to effectively reflect the health level of individuals (Fayers and Sprangers, 2002) . The questionnaire used in this study mainly includes two parts: personal basic attributes and self-rated health scale. The basic personal attributes mainly include 10 items, including sample number, gender, age, occupation, marital status, address, length of residence, residential area, number of families, agricultural labor situation ,and annual family income. SRHMS v1.0 (Self-rated Health Measurement Scale V1.0) was used in the self-rated health scale. SRHMS v1.0 has good content validity, structure validity and calibration validity, which is more in line with the characteristics and habits of Chinese residents, and is suitable for all kinds of people over the age of 14. The content of the scale consists of three subscales of self-rated physical health, mental health and social health, as well as four parts of the overall self-evaluation, with a total of 48 items, which can be divided into nine dimensions: physical symptoms and organ function (B1), daily physical activity (B2), physical activity ability (B3), psychosocial symptoms and negative emotions (M1), positive emotions (M2), cognitive function (M3), role activity and social adaptability (S1), social resources and social contact (S2) and social support (S3). The overall self-assessment includes the overall physical self-test, the overall psychological self-test, the overall social self-test and the overall health self-test. The maximum score of each item is 10 points, the minimum is 0 points, and the maximum possible score is 440 points after removing the self-evaluation part of auxiliary verification. The scores of items 4, 5, 7, 24, 25, 26, 27, 28, 29 and 30 are scored in reverse to ensure the validity of the questionnaire (Xu *et al.*, 2000; Xu, 2001; Chuqun *et al.*, 2021)Δ

Due to potential problems such as residents' personal willingness to fill in, only 125 valid health questionnaire data were collected in this study. Table 3 shows the basic information data of the respondents, and Fig 3 shows the distribution of the total self-rated health scores of the sample population.

Table 3. General Information of the Respondents

The General Information of the Respondents(N=125)

Item		Number of samples(%)
Village name		
	Hongfu Community	19(15.2%)
	Dongyue Community	30(24%)
	Shuangshi Community	26(20.8%)
	Bajiaojing Village	3(2.4%)
	Fuxing Community	15(12%)
	Mingyue Village	32(25.6%)
Gender		
	male	52(41.6%)
	female	73(58.4%)
Age		
	Children and adolescents (0-19)	5(4%)
	Young adults (20-39)	35(28%)
	Middle aged adults (40-59)	60(48%)
	Elderly (≥ 60)	25(20%)
occupation		
	farmer	102(81.6%)
	Public officials	1(0.8%)
	Individual business	2(1.6%)
	worker	2(1.6%)
	Others (including unfilled)	18(14.4%)
Living conditions		
	Local resident population	106(84.8%)
	Immigrant population	9(7.2%)
	Others (including unfilled)	10(8%)
Residence		
	Old house	25(20%)
	New rural community	94(75.2%)
	Houses in counties and other cities	1(0.8%)
	Others (including unfilled)	5(4%)
Years of residence		
	≤ 10 years	29(23.2%)
	11-30 years	30(24%)
	31-50 years	33(26.4%)
	≥ 51 years	30(24%)
	Others (including unfilled)	3(2.4%)
marital status		
	unmarried	11(8.8%)
	married	104(83.2%)

	Divorce	4(3.2%)
	Bereavement	4(3.2%)
	Others (including unfilled)	2(1.6%)
Family size		
	<3 persons	7(5.6%)
	3-5 persons	86(68.8%)
	□ 5 persons	24(19.2%)
	Others (including unfilled)	8(6.4%)
Annual household income		
	<10000	15(10.44%)
	10000-20000	18(17.4%)
	20000-30000	35(25.0%)
	30000-50000	18(14.6%)
	50000-100000	22(20.1%)
	100000-200000	5(3.5%)
	>200000	2(1.4%)
	Others (including unfilled)	10(7.6%)
Farming conditions		
	Yes, there is still farmland to be cultivated	107(85.4%)
	No. no more farmland	18(13.9%)

Note: processed according to the questionnaire data

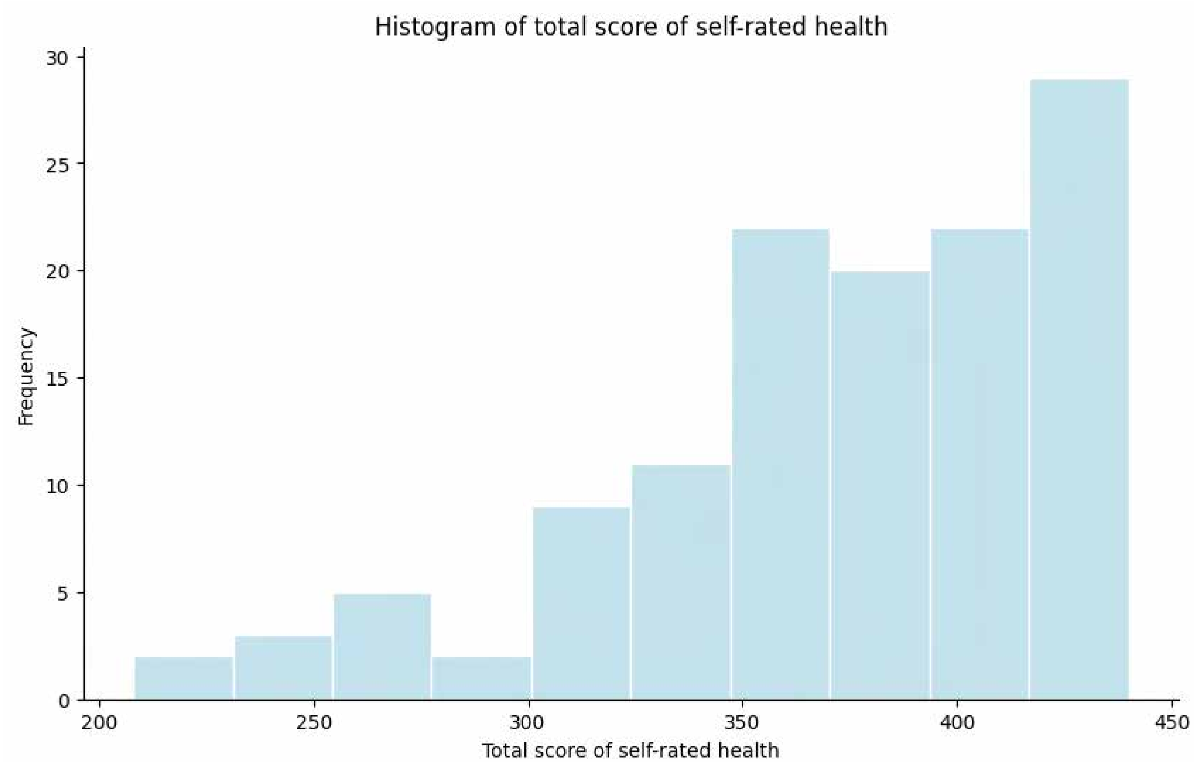

Fig 3. Histogram of total score of self-rated health

Built environment data based on Web Map

The built environment data used in this study mainly include POI data and building vector contour data within the scope of the study. POI data is obtained from Amap, which includes 22 categories and is limited within the administrative boundaries of Pujiang County, Lushan County, Mingshan District of Ya'an City and Qionglai City, with a total of 52784 items. In previous studies, the original POI data is usually reclassified according to the research purpose, such as taking the residents' service attribute as the core (Yang et al., 2021) and the urban functional area attribute as the core (Gu et al., 2018; Yang et al., 2020) . In order to meet the requirements for the identification of stay activity functional areas in this study, the POI of Amap is reclassified into 11 categories (Table 4) by focusing on the attributes of resident services, so as to facilitate the identification of subsequent residents' travel characteristics. The building contour vector data is obtained from the MapWorld, which mainly contains the outer contour vector shape data of all buildings in the three counties and cities, so as to help judge the resident stay activity category in the POI missing area.

Table 4. POI data category reclassification

POI data category reclassification

Amap category	Category after reclassification	Simple category name
Enterprises (Factory, Farming, Forestry, Animal Husbandry and Fishery Base)	Industrial and agricultural production	Labor
Enterprises (Enterprises, Famous Enterprise, Company)	Enterprises and offices	Work
Finance & Insurance Service	Financial service	Finance
Science/Culture & Education Service	Educational and cultural services	Education
ResidenceResidence and related servicesPlace Name & Address (Village name, building number, community name and doorplate information)		
Commercial House		
Accommodation Service		
MedicalMedical related services-Medical Service		
EntertainmentLeisure and entertainment servicesSports & Recreation		

TourismRecreational tourismPlace Name & Address(Natural Place Name)

Tourist Attraction

TransportationTransportation facilities and related servicesTransportation Service

Auto Service

Auto Dealers

Auto Repair

Motorcycle Service

ShoppingCommercialand life servicesFood & Beverages

Daily Life Service

Shopping

MunicipalGovernment and public servicesGovernmental Organization & Social Group

Public Facility

Methods

Travel trajectory activity point acquisition

At present, there are four main directions for the research of stay activity point recognition in travel trajectory. First, based on the fitting of road network information and travel trajectory, the distance between trajectory points and road vectors is calculated to identify stay points (Du and Aultman-Hall, 2007). This method needs more detailed regional road network information as an auxiliary, which is suitable for use in urban areas with relatively complete road network data. It is not suitable for use in rural natural environment with insufficient construction of standard road network and serious lack of road network data. The second is to use the velocity characteristics of the trajectory points to judge and detect the trajectory points whose velocity approaches zero as the stay points (Zhang et al., 2013) . This method is aimed at the stay points in slow travel activities in rural areas, and requires equipment with high-precision speed collection ability. The third is to judge the stay point based on geographical semantics. For example, the SMoT (Stops and Moves of Trajectories) model first presets the locations of the points of interest that the samples often go to, generates appropriate geographic geometric entities based on such locations, and judges the stay activities through the residence characteristics of the sample motion trajectory in the geometric entities (Alvares et al., 2007). The fourth is based on density clustering to identify stay points. As the stay activities will be concentrated in a region for a long time, the stay point can be better found through density judgment. DBSCAN algorithm based on density clustering has a good performance in the identification of travel purpose. It only needs the location and time stamp data of trajectory points. By setting the clustering limit parameters such as EPS neighborhood and Minpts, it can expand the core points, divide the clusters by density reachability, and calculate the cluster center point as the stay point (Ester et al., 1996) .

On this basis, Zhou Yang and Yang Chao analyzed the shortcomings of existing ST-DBSCAN (Birant and Kut, 2007), TrigDBSCAN(Yan, 2011), C-DBSCAN(Gong et al., 2015) algorithms in the required data complexity, scientific threshold, core point search method, cluster merging method, and path overlap, and proposed the AT-DBSCAN algorithm (Zhou and Yang, 2018) The algorithm checks the internal trajectory points by setting a fixed length sliding window, so as to cluster and identify the core points, forming the initial cluster and cluster, and finally calculates the center points as the stay points after filtering the cluster according to the length of time. It requires the initial data and can better eliminate the influence of noise in the original trajectory data.

Due to the difficulty of data collection in rural areas, where residents' activities involve non built environments such as farmland and wilderness, it is difficult to manually determine the range of interest locations. In addition, the GPS data collected in this study has certain errors and noise effects. Therefore, this paper chooses to draw on the idea of AT-DBSCAN algorithm for identifying travel trajectory stay points. The AT-DBSCAN algorithm has five main parameters that need to be set, namely the sliding window length "k", time interval "I", spatial neighborhood distance "Eps", minimum trajectory point count threshold "MinPts", and minimum stay points duration "DU". Due to the fact that the GPS trajectory points data recording interval used in this study is 1 minute, considering that stay activities with a duration of too short are of little significance to this study, stay activity points with a duration of more than 15 minutes are included in the research scope. In this study, k=21, I=10min, Eps=60m, MinPts=10, DU=15min are set, and 172 residents' trajectories are respectively clustered to identify the stay activity points. Firstly, use a sliding window to mark trajectory points within the spatial neighborhood with a number of neighboring points exceeding the MinPts value as core points. Then, starting from random core points, traverse all unclustered core points and cluster the core points within the spatial neighborhood Eps range with a time interval of no more than 2 minutes as the initial cluster. Then start from the initial cluster with the highest density, merge other adjacent clusters with a time interval less than I, repeat this operation until all the clusters that can be merged are merged, and finally exclude the clusters with a duration less than Du. The final data includes the longitude and latitude coordinates of the stay activity point, the cluster number, the start and end time of the stay activity, the stay time, the date number, and the sample number (Table 5). An example of space-time diagram of clustered data trajectories is shown in Fig 4.

Table 5. Stay point information

Stay point information							
Longitude	Latitude	Cluster number	Start time	End time	Stay time (minutes)	Date	ID
103.4714407	30.4045981	1	08:16:00	08:32:00	16	12	9618007128
103.4600468	30.400223	2	08:55:00	10:03:00	68	12	9618007128
103.4602874	30.40206229	3	10:11:00	10:27:00	16	12	9618007128
103.3760856	30.27748183	4	12:17:00	18:51:00	394	12	9618007128
103.3756631	30.27752271	5	00:12:00	06:42:00	390	12	9618007128
103.4726061	30.40617506	6	07:42:00	08:11:00	29	12	9618007128
103.4604299	30.40223265	7	10:30:00	10:56:00	26	12	9618007128
…	…	…	…	…	…	…	…

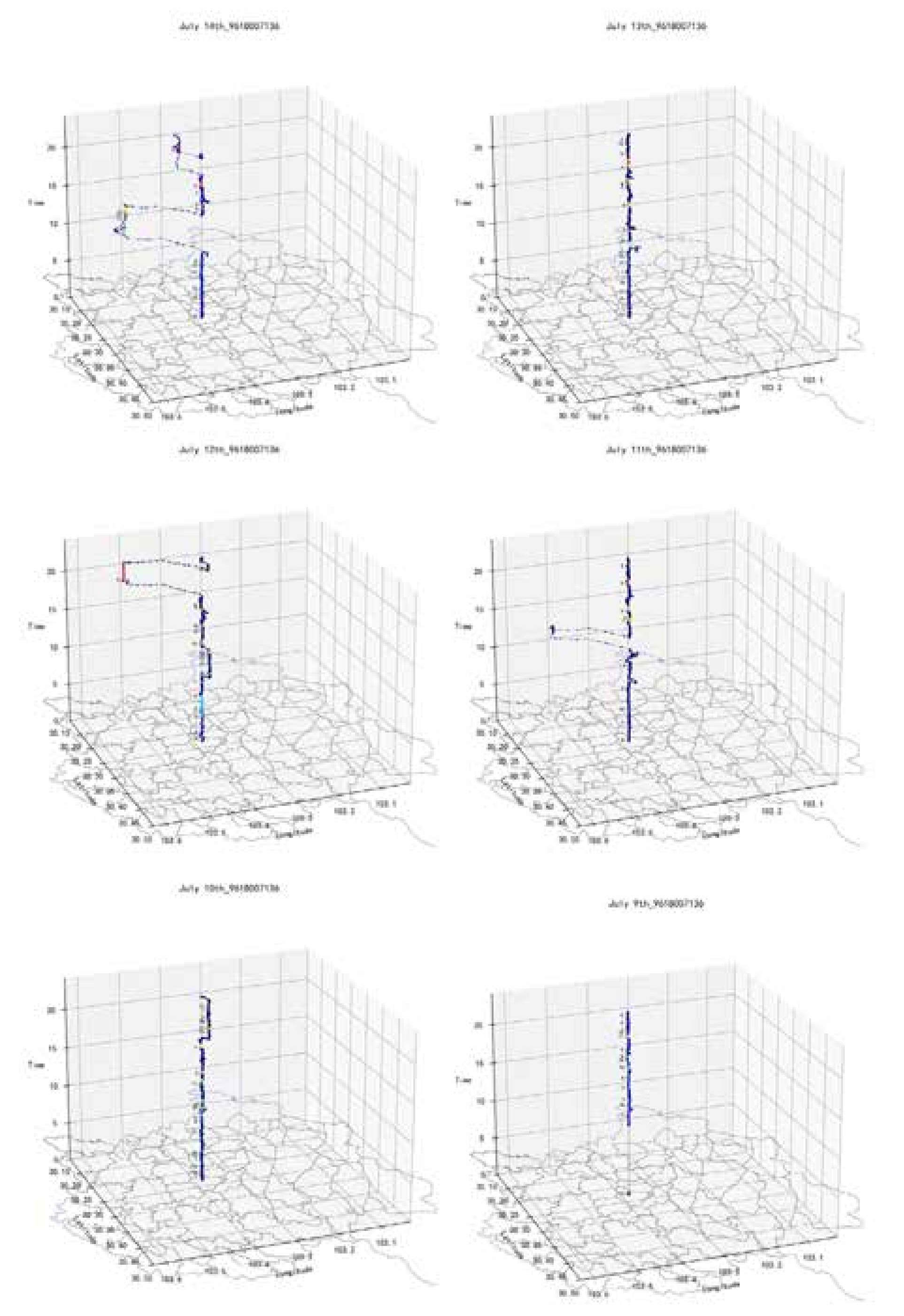

Fig 4. Example of Trajectory space-time diagram

Stay activity attribute judgment

After obtaining the stay activity point of the sample by clustering, it is necessary to infer its functional type. We choose to use POI data to judge, refer to the combination of TF-IDF algorithm and grid division method used in the related research of function area judgment, identify POI weights, determine grid function attributes, and further infer the type of stay activity points (Sparck Jones, 1972; Zhu, 2024). The specific method is to divide the area within the research scope into multiple grids according to the side length of 120m, and determine the weight of each functional category in the POI data falling into the grid. In the weight calculation part, using the good classification ability of TF-IDF algorithm, each grid area is regarded as a single text, and the proportion of POI categories in a single text is regarded as word frequency. IDF (Inverse Document Frequency) is used to represent the importance of different POI categories, and finally the POI weights of each category in each grid are obtained. Then the weight proportion is calculated, and the function of grid area is divided according to the proportion. In this study, when the weight of a certain type of POI in the grid area accounts for more than 50%, it is considered that the functional category of the grid area is the same. When the weight of two types of POI is between 20% and 50%, it is considered that the function category of the grid area is a combination of two types of functions. When the weight of all POIs is 0%, the function category of the grid area is considered as no data area. Other cases are classified as comprehensive functions. After determining the function of the grid area, assign it to the function type of all stay activity points falling in the grid.

However, in rural areas, there are many grids that are judged as no data area due to the lack of POI. In fact, these areas include built-up environmental areas such as villages and settlements, as well as natural areas such as fields and mountains. Therefore, it is necessary to combine the building contour vector data to assist the judgment. In rural areas, areas without POI but with buildings are generally within villages where residents live, and these areas can be regarded as functional areas of “Residence”. In this study, for the trajectory stay point marked as “no data area”, judge whether there is a building outline intersecting with it within the 120m diameter of the section. If there is, change the functional attribute of the point to “Residence”, otherwise, change it to “Natural area”. Finally, the potential activity type characteristics of each stay activity points are obtained.

Travel activity chain clustering

After obtaining the functional attributes of the sample stay activity points, it is necessary to build the sample daily activity chain. We first divide the trajectory of each sample in the time range of 8:00-10:00 into 28 time blocks in 30 minutes. Then judge whether the time range of each time block is within the start and end time range of the stop activity point in order. If it is within the range, the function attribute of the stay active point is assigned to the time block. Then all time blocks that are not assigned the function attribute of the active point are identified as “moving” activities. Finally, the activity chain data with granularity of 30 minutes is formed. Due to the relatively small sample size of 124 data collected by the self-rated health questionnaire, in order to ensure the consistency between the subsequent analysis and the self-rated health questionnaire data, the activity chain without corresponding questionnaire data was deleted according to the sample number before the activity chain clustering, and 709 travel activity chain data remained.

In the part of activity chain clustering, this paper draws on the trajectory vectorization and clustering method based on word vector model (word2vec) proposed by liwenxiang et al. (Li et al., 2024) In this study, the activity unit is mapped to a word vector, and the activity chain is mapped to a sentence vector, so that the activity pattern recognition problem is transformed into the semantic similarity measurement between the sentence vectors of the activity chain. Through cluster analysis, user groups with similar activity patterns can be identified. This method shows good effect in multi day activity pattern recognition. The daily activity chain data of all samples are converted into comma separated sentence text composed of functional attributes such as “Residence”, “Labor”, “Work”, “Transportation”, “Finance”, “Education”, “Medical”, “Tourism”, “Entertainment” “Shopping”, “Municipal”, “Comprehensive”, “Natural”, “Move”, in chronological order. Then use text data to train word2vec model, set the word vector dimension vector_size=19, the context window size window=20, and the minimum occurrence frequency of reserved words min_count=1, that is, words that appear at least once in the corpus will be retained. Finally, the daily travel activity chain sentence vector of each sample is generated, and SVD (Singular Value Decomposition) method is used to decompose the sentence vector to remove the main components in the data, so as to achieve normalization, which is convenient for feature recognition in subsequent clustering(Gargiulo et al., n.d.)Δ

After that, this study chose to use k-means algorithm to cluster the sentence vectors generated by the daily activity chain of samples. The algorithm has good interpretability for the results. The centroid of each cluster is the mean value of all data points in the cluster, which can directly reflect the central position and main characteristics of the cluster data (Li et al., 2020; Lin et al., 2009) . Because the k-means algorithm needs to specify the number of clusters in advance, the elbow method combined with the comprehensive judgment of contour coefficient can better distinguish the appropriate number of clusters (Humaira and Rasyidah, 2020), as shown in Fig 5. Combined with the actual travel activities, elbow diagram and contour coefficient results, the preset clustering number k=7 is selected to obtain the final clustering data.

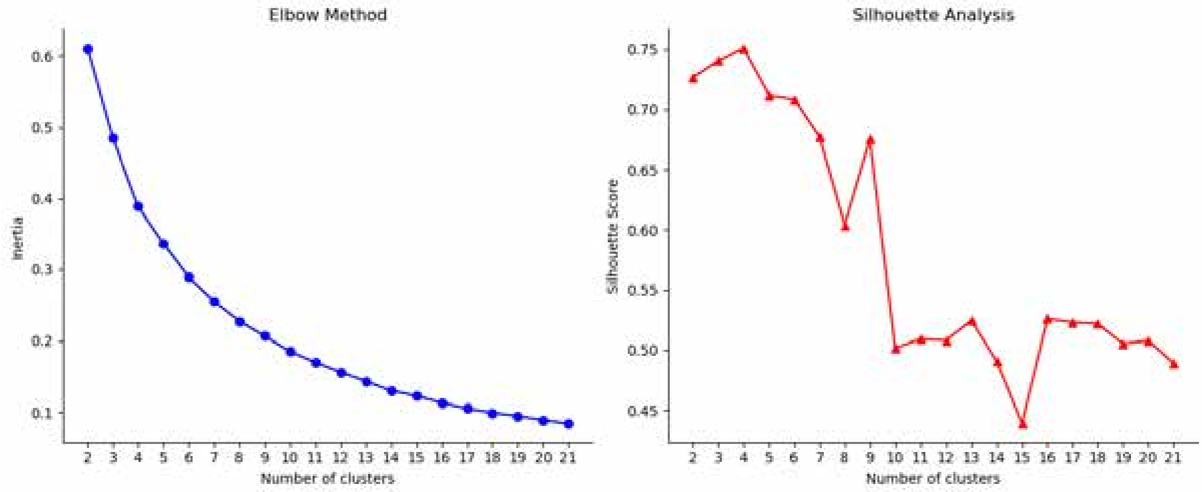

Fig 5. Elbow diagram and contour coefficient diagram Fig 6.

Analysis of health level differences among different travel mode
After mapping the self-rated health questionnaire data with the clustered activity chain data according to the sample number, it is necessary to analyze the difference of health level between different activity chain clusters. The commonly used statistical analysis methods for differences mainly include Analysis of Variance, Chi-Square Test, and Nonparametric test (le Cessie et al., 2020; Lee et al., 2013). Variance analysis requires that the dependent variable is a continuous variable and meets the homogeneity and normality of variance. Chi square test is applicable to the case where the dependent variable is classified variable. Using SPSSAU data analysis platform, the health score data corresponding to all travel activity chains were normalized and then tested for normality(SPSSAU, 2025). The results showed that the p value was less than 0.05, and the residents' self-rated health score data did not conform to the normal distribution (Table 6). Since the distribution of self-rated health data does not conform to the normal distribution, and the dependent variable is a continuous variable, this study uses a nonparametric test to analyze the difference of self-rated health scores between clusters in the activity chain.

Table 6. Analysis results of normality test

Analysis results of normality test

Name	sample size	average value	standard deviation	skewness	kurtosis	Kolmogorov-Smirnov		Shapiro-Wilk	
						Statistic D	p	Statistic W	p
Total health score(normalization)	650	0.715	0.228	-0.867	0.345	0.106	0.000**	0.928	0.000**

* *p*<0.05 ** *p*<0.01

Results
In order to better analyze the activity patterns represented by each activity chain cluster, the functional attribute weights of all activity chains in each cluster are calculated and represented as images (Fig 6). It can be seen that the stay activities of the samples in cluster 0 are distributed in "Natural", "Comprehensive", "Residence" and "Shopping", and the weight of "Residence" is relatively high, but not more than 50%. The stay activities in cluster

1 are obviously concentrated in “Residence”, and a small part are distributed in “Natural”. The stay activities in cluster 2 are less distributed in “Residence “, relatively balanced in “Finance”, “Medical” and “Shopping”, and there are stay activities in some “Medical” areas. In cluster 3, “Medical” has the highest weight, followed by “Residence” and “Shopping”, and there are also a few stay activities in the “Medical” area. The stay activities in Cluster 4 are obviously concentrated in the “Comprehensive” area. The stay activities in cluster 5 are mainly “Work”. Cluster 6 is mainly distributed in “Natural”, “Tourism”, “Entertainment”, “Shopping” and other “Entertainment” areas, but in addition to “Residence”, there are also “Medical” areas in other activities.

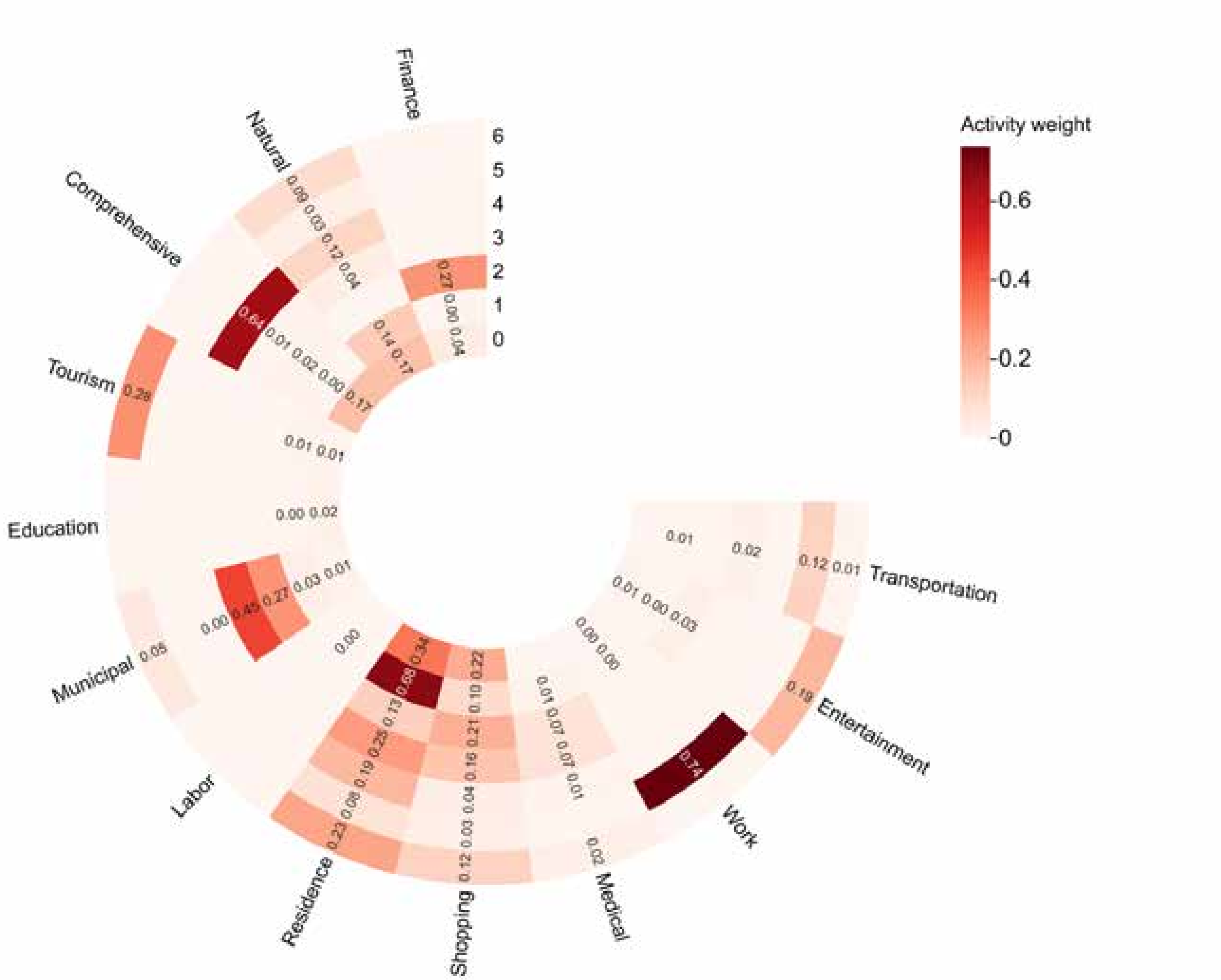

Fig 6. Feature weight of clustering activities

Nonparametric test was used to analyze the health differences among clusters of activity patterns. The results are shown in the table below (Table 7). The original hypothesis was that there was no difference in the self-rated health scores among the activity mode clusters. Since the p value was equal to 0.011 and less than 0.05, the original hypothesis was rejected and it was considered that there were differences in the self-rated health scores among the activity mode clusters. Draw a box plot to further analyze the specific differences in self-rated health scores among various activity modes (Fig 7).

The median of self-rated health score in cluster 2 is the highest 440, and the data distribution is relatively concentrated, with small variance, but the sample size is too small, and it shows an obvious left deviation distribution. The median of self-rated health score in cluster 0 was 403, which was at the second high level. 75% of the data were concentrated between 328 and 419.8, which had a certain concentration. The median self-rated health score of cluster 1 was 382.0, and the score distribution was between 339.5 and 414.0. Compared with cluster 0, the 75% data distribution was more concentrated. The median self-rated health score of cluster 3 was 373.0, and 75% of the scores were distributed between 271.0 and 420.0, indicating that the cluster had a wide range of health scores, and the self-rated health scores were not concentrated enough and showed a left skewed distribution, with more head data. The median self-rated health score of cluster 6 was 356.0, and 75% of the scores were concentrated between 294.0 and 385.0, which was significantly lower than that of cluster 0, 1, 2 and 4. The self-rated health score data of cluster 5 are all 352. Due to the small amount of data, it is considered as a special activity chain cluster with obvious characteristics. The median self-rated health score of Cluster 4 was 348.5, and 75% of the scores were distributed between 323.0 and 422.0, showing a right deviation distribution, with more tail data. Combined with the median and data distribution, the overall self-rated health score in this cluster is lower than that in clusters 0, 1, 2 and 3.

Table 7. Nonparametric test analysis results

Nonparametric test analysis results

*p*Kruskal-Wallis test statistic H Cluster median M(P_{25} P_{75})
6.0(*n*=19)5.0(*n*=5)4.0(*n*=6)3.0(n=19)2.0(*n*=6)1.0(n=626)0.0(n=28)
0.011*16.477356.0
(294.0,385.0)352.0
(352.0,352.0)348.5
(323.0,422.0)373.0
(271.0,420.0)440.0
(398.8,440.0)382.0
(339.5,414.0)403.0
(328.0,419.8)Total score of self-rated health

* *p*<0.05 ** *p*<0.01

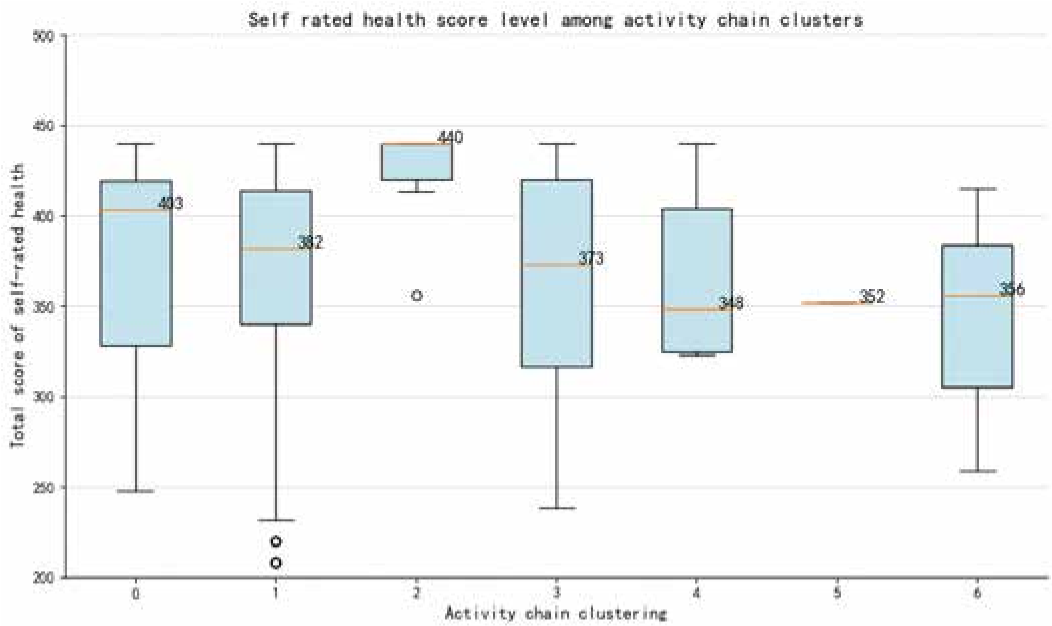

Fig 7. Box plot of self-rated health scores among different activity clusters

Discussion
This study found that in the daily life of residents in rural areas, the scope of activities of most people is relatively small, only staying in the village and its surrounding areas. According to the actual situation, cluster 0 is defined as "residential-mixed" travel activity mode, cluster 1 as "residential-natural" travel activity mode, cluster 2 as "commercial service" travel activity mode, cluster 3 as "municipal facilities" travel activity mode, Cluster 4 as "comprehensive functional" travel mode, cluster 5 as "office" travel mode, and cluster 6 as "entertainment" travel mode.

The health level of "commercial service" travel activity mode is the highest, and its activities are mainly concentrated in financial, municipal, commercial services and other areas, and a certain proportion of activities occur in medical areas. Commercial activities are usually associated with higher economic income and better living conditions, which may further promote the health status of residents (Contoyannis and Jones, 2004). However, the data volume of this cluster is small, and the data presents a significant left biased distribution, which may limit the universality of the results. Future studies need a larger sample size to verify this finding.

The self-rated health score of the "residential-mixed" travel activity mode is the second highest, covering activities in a variety of functional areas, including residential and related services, comprehensive functional areas, commercial service areas, etc. This diversified mode of travel activities may enable residents to access different types of resources and services, thus having a comprehensive positive impact on health. Residential areas usually provide basic living security, while comprehensive functional areas and business service areas may provide additional facilities and social opportunities. These factors work together to help maintain a high level of health.

The "residential-natural" travel activity mode has the largest amount of data, and the overall health level is widely distributed, and most of them are concentrated in the high score level, second only to "commercial service" and "residential-mixed". The stay activity of this mode are mainly concentrated in "Residence" and "Natural". According to the questionnaire, most rural residents need to go to farmland and woodland for labor in their daily life, and the moderate intensity physical activity brought by labor may promote their health level (Blair et al., 1992; Posadzki et al., 2020) . At the same time, the natural environment usually has a positive impact on mental and physical health, helping to relieve stress, improve mood and enhance physical activity (Ewert and Chang, 2018; Jimenez et al., 2021). However, there are still some residents with relatively low scores in this mode. This may be related to the quality of the living environment, the accessibility and utilization of natural areas, and the sensitivity of individuals to the natural environment. In addition, the relatively single type of activities may limit the healthy development of residents to a certain extent. The sample size of this cluster is 626 samples, which is one of the clusters with the largest sample size. The results have certain representativeness and reliability, but the impact of individual differences on health still needs to be considered.

The stay activities of the “municipal facility” travel activity mode are mainly concentrated in the “Municipal”, “Residence” and “Shopping” areas, and there are also a few stay activities in the “Medical” areas. The use of municipal service facilities provides residents with basic public services and convenience, and helps to promote the healthy lifestyle of residents. However, the cluster has a wide range of health scores, and the median advantage is not prominent, which may indicate that different individuals have different health benefits in the use of municipal facilities, which are affected by many factors such as individual health awareness, lifestyle, facility accessibility and so on. And the amount of clustering data is small, and the data distribution may also be affected by the sample characteristics. Further research is needed to clarify the specific mechanism behind it.

The “comprehensive functional” travel activity mode is mainly parked in the comprehensive functional area, and the overall level of self-rated health score is lower than the first few. Comprehensive functional areas usually integrate many functions such as residence, commerce, office and leisure. Although they provide convenient living and working conditions, such areas may also have adverse health factors, such as noisy environment, poor quality and high density, which will have a negative impact on the health of residents, and then affect their health. The amount of clustering data is small, and the data distribution may have large deviation. More research is needed to verify this hypothesis.

The stay of “leisure and entertainment” travel activity mode is mainly distributed in “Natural”, “Tourism”, “Entertainment”, “Shopping” and other areas, as well as “Residence” and “Medical” areas. Leisure and recreational activities are generally considered to be beneficial to mental and physical health, and can relieve stress, enhance physical activity and promote social interaction (Fancourt D et al., 2021; Paggi et al., 2016). However, the health score of this cluster is not high, which may be related to the type, duration, frequency of leisure activities and the subjective experience and response of individuals to these activities, and may also be related to the stay activity of these data in the “medical” area. The amount of clustering data is small, and the data distribution has some limitations. More research is needed to explore the complex relationship between leisure and entertainment activities and health.

The self-rated health score of the “office” travel activity mode is the lowest, mainly in the “office” area. Long term office activities may lead to problems such as reduced physical activity, increased work pressure, and increased psychological burden, which may have a negative impact on health (Kouvonen et al., 2013) . However, the sample size is small and the source is single, which may be due to the small number of people in the daily regular office travel mode in rural areas.

Conclusion

The results show that there are differences in self-rated health level among rural community residents with different travel activity modes. Among them, the residents’ health score of the “commercial service” model is the highest. It is speculated that social interaction and economic interaction in business activities have a positive effect on health, and may also be related to the higher economic level of people engaged in business activities; The health level of residents in the “residential-mixed” mode is second, and the diversified types of activities enable residents to contact multiple types of resources and services, which has a certain promoting effect on health. The “residential-natural” travel activity mode is the main mode of daily travel for residents in rural areas. In this type of travel activity mode, most residents’ self-rated health level is relatively high, which may be related to the fact that some residents in rural areas are engaged in agricultural activities to promote physical activity. The residents’ self-rated health score level of the “municipal facility” travel activity mode is relatively scattered, indicating that the travel mode has no obvious centralized health level characteristics, and the health level of the group may have more complex correlation with personal factors. The residents’ self-rated health score of the “comprehensive functional” travel activity mode is lower than that of the first three activity modes. It is speculated that the comprehensive functional areas in rural areas may have poor environmental quality, which is not conducive to health promotion. The self-rated health score of residents in the “entertainment” mode is relatively poor, which is contrary to the generally believed promotion effect of leisure and entertainment activities on health. This kind of activity mode has the activity stay in the medical area, which may indicate that these people's health level is poor, so they choose leisure and entertainment activities for treatment, leading to different conclusions. The health score of residents in "office" mode is the lowest, which is consistent with the conclusion that physical activity decreases and stress increases due to long-time office work in previous studies. In view of the differences between different travel activity modes and the self-rated health level of rural community residents, it is suggested to reasonably arrange various functional areas in rural community planning, encourage the development of commercial services and mixed functional areas, enhance economic vitality and social communication opportunities, and promote residents' health. In rural areas, the environmental quality of comprehensive functional areas should be optimized to avoid adverse effects on health. At the same time, increasing the planning and configuration of leisure areas should not only meet the residents' needs for treatment, but also consider how to enhance the health promotion effect of activities. In addition, when planning the office area, we should pay attention to the health status of office residents, reasonably arrange the working hours and environment, and reduce the health risks caused by reduced physical activity and increased stress, so as to comprehensively improve the health level of rural community residents.

However, this study also has limitations. The sample size is relatively small and concentrated in specific areas, which may affect the universality of the results; The data collection time is short and does not cover the impact of different seasonal changes on travel and health. Future research should expand the sample size and geographical scope of rural areas, extend the time of data collection, and combine more objective health indicators with travel monitoring methods to further verify and deepen the research conclusions, so as to provide more comprehensive and accurate guidance for rural community health planning.

References

1. Aerts, R., Honnay, O. and Van Nieuwenhuyse, A. (2018), "Biodiversity and human health: mechanisms and evidence of the positive health effects of diversity in nature and green spaces", *British Medical Bulletin*, Vol. 127 No. 1, pp. 5–22, doi: 10.1093/bmb/ldy021.

2. Akinlotan, M., Primm, K., Khodakarami, N., Bolin, J. and Ferdinand, A.O. (2021), "Rural-Urban Variation in Travel Burden for Care: Findings from the 2017 National Household Travel Survey".

3. Ali, M., Dharmowijoyo, D.B.E., Harahap, I.S.H., Puri, A. and Tanjung, L.E. (2020), "Travel behaviour and health: Interaction of Activity-Travel Pattern, Travel Parameter and Physical Intensity", *Solid State Technology*, Vol. 63 No.

4. Ali, M., Macioszek, E. and Yuen, C.W. (2024), "Health enhancement through activity travel participation and physical activity intensity", *Journal of Transport & Health*, Vol. 39, p. 101927, doi: 10.1016/j.jth.2024.101927.

5. Alvares, L.O., Bogorny, V., Kuijpers, B., de Macedo, J.A.F., Moelans, B. and Vaisman, A. (2007), "A model for enriching trajectories with semantic geographical information", *Proceedings of the 15th Annual ACM International Symposium on Advances in Geographic Information Systems*, Association for Computing Machinery, New York, NY, USA, pp. 1–8, doi: 10.1145/1341012.1341041.

6. Berg, J. and Ihlström, J. (2019), "The Importance of Public Transport for Mobility and Everyday Activities among Rural Residents", *Social Sciences*, Multidisciplinary Digital Publishing Institute, Vol. 8 No. 2, p. 58, doi: 10.3390/socsci8020058.

7. Birant, D. and Kut, A. (2007), "ST-DBSCAN: An algorithm for clustering spatial–temporal data", *Data & Knowledge Engineering*, Vol. 60 No. 1, pp. 208–221, doi: 10.1016/j.datak.2006.01.013.

8. Blair, S.N., Kohl, H.W., Gordon, N.F. and Jr, R.S.P. (1992), "How Much Physical Activity is Good for Health?", *Annual Review of Public Health*, Annual Reviews, Vol. 13 No. Volume 13, 1992, pp. 99–126, doi: 10.1146/annurev.pu.13.050192.000531.

9. Bratman, G.N., Anderson, C.B., Berman, M.G., Cochran, B., Vries, S. de, Flanders, J., Folke, C., *et al.* (2019), "Nature and mental health: An ecosystem service perspective", *Science Advances*, American Association for the Advancement of Science, doi: 10.1126/sciadv.aax0903.

10. Briggs, D.J., De Hoogh, K., Morris, C. and Gulliver, J. (2008), "Effects of travel mode on exposures to particulate air pollution", *Environment International*, Vol. 34 No. 1, pp. 12–22, doi: 10.1016/j.envint.2007.06.011.

11. le Cessie, S., Goeman, J.J. and Dekkers, O.M. (2020), "Who is afraid of non-normal data? Choosing between parametric and non-parametric tests", *European Journal of Endocrinology*, Oxford Academic, Vol. 182 No. 2, pp. E1–E3, doi: 10.1530/EJE-19-0922.

12. Chuqun, G.A.O., Hechuan, Z., Songlan, L.U.O., Yajing, Z., Liangji, G.U.O. and Lili, Z. (2021), "Applicability of the SRHMS V1.0 in the Rural Elderly∆an Analysis of 2 048 Cases", *Chinese General Practice*, Vol. 24 No. 19, p. 2418, doi: 10.12114/j.issn.1007-9572.2021.00.218.

13. Contoyannis, P. and Jones, A.M. (2004), "Socio-economic status, health and lifestyle", *Journal of Health Economics*, Vol. 23 No. 5, pp. 965–995, doi: 10.1016/j.jhealeco.2004.02.001.

14. Dėdelė, A., Miškinytė, A., Andrušaitytė, S. and Nemaniūtė-Gužienė, J. (2020), "Dependence between travel distance, individual socioeconomic and health-related characteristics, and the choice of the travel mode: a cross-sectional study for Kaunas, Lithuania", *Journal of Transport Geography*, Vol. 86, p. 102762, doi: 10.1016/j.jtrangeo.2020.102762.

15. Du, J. and Aultman-Hall, L. (2007), "Increasing the accuracy of trip rate information from passive multi-day GPS travel datasets: Automatic trip end identification issues", *Transportation Research Part A: Policy and Practice*, Vol. 41 No. 3, pp. 220–232, doi: 10.1016/j.tra.2006.05.001.

16. Ester, M., Kriegel, H.-P. and Xu, X. (1996), "A Density-Based Algorithm for Discovering Clusters in Large Spatial Databases with Noise", Vol. 96(34), presented at the kdd, pp. 226–231.

17. Ewert, A. and Chang, Y. (2018), "Levels of Nature and Stress Response", *Behavioral Sciences*, Multidisciplinary Digital Publishing Institute, Vol. 8 No. 5, p. 49, doi: 10.3390/bs8050049.

18. Fancourt D, Aughterson H, Finn S, and Emma W. (2021), "How leisure activities affect health: a narrative review and multi-level theoretical framework of mechanisms of action", *The Lancet Psychiatry*, Elsevier, Vol. 8 No. 4, pp. 329–339, doi: 10.1016/S2215-0366(20)30384-9.

19. Fayers, P.M. and Sprangers, M.A. (2002), "Understanding self-rated health", *The Lancet*, Elsevier, Vol. 359 No. 9302, pp. 187–188, doi: 10.1016/S0140-6736(02)07466-4.

20. Foley, L., Dumuid, D., Atkin, A.J., Olds, T. and Ogilvie, D. (2018), "Patterns of health behaviour associated with active travel: a compositional data analysis", *International Journal of Behavioral Nutrition and Physical Activity*, Vol. 15 No. 1, p. 26, doi: 10.1186/s12966-018-0662-8.

21. Frank, L.D., Iroz-Elardo, N., MacLeod, K.E. and Hong, A. (2019), "Pathways from built environment to health: A conceptual framework linking behavior and exposure-based impacts", *Journal of Transport & Health*, Vol. 12, pp. 319–335, doi: 10.1016/j.jth.2018.11.008.

22. Gao D., Shi L., Lin J., Wang X. and Wang D. (2024), "Research on efficiency and equity of rural medical and health resources allocation in China based on DEA-GIS method", *Chinese General Practice*, Vol. 27 No. 7, pp. 849–856.

23. Gargiulo, A.A., Crisostomi, D., Bucarelli, M.S., Scardapane, S., Silvestri, F. and Rodola, E. (n.d.). "Task Singular Vectors: Reducing Task Interference in Model Merging".

24. Gong, L., Sato, H., Yamamoto, T., Miwa, T. and Morikawa, T. (2015), "Identification of activity stop locations in GPS trajectories by density-based clustering method combined with support vector machines", *Journal of Modern Transportation*, Vol. 23 No. 3, pp. 202–213, doi: 10.1007/s40534-015-0079-x.

25. Gu Y., Jiao L., Dong T., Wang Y. and Xu G. (2018), "Identification and interaction analysis of urban functional areas based on multi-source data", *Geomatics and Information Science of Wuhan University*, No. 7, pp. 1113–1121, doi: 10.13203/j.whugis20160192.

26. Hearn, M., Pinto, C. and Moss, J.L. (2024), "Evaluating the Connection Between Rural Travel Time and Health: A Cross-Sectional Analysis of Older Adults Living in the Northeast United States", *Journal of Primary Care & Community Health*, SAGE Publications Inc, Vol. 15, p. 21501319241266114, doi: 10.1177/21501319241266114.

27. Humaira, H. and Rasyidah, R. (2020), "Determining The Appropiate Cluster Number Using Elbow Method for K-Means Algorithm", *Proceedings of the Proceedings of the 2nd Workshop on Multidisciplinary and Applications (WMA) 2018, 24-25 January 2018, Padang, Indonesia*, presented at the Proceedings of the 2nd Workshop on Multidisciplinary and Applications (WMA) 2018, 24-25 January 2018, Padang, Indonesia, EAI, Padang, Indonesia, doi: 10.4108/eai.24-1-2018.2292388.

28. Jimenez, M.P., DeVille, N.V., Elliott, E.G., Schiff, J.E., Wilt, G.E., Hart, J.E. and James, P. (2021), "Associations between Nature Exposure and Health: A Review of the Evidence", *International Journal of Environmental Research and Public Health*, Multidisciplinary Digital Publishing Institute, Vol. 18 No. 9, p. 4790, doi: 10.3390/ijerph18094790.

29. Kouvonen, A., Vahtera, J., Oksanen, T., Pentti, J., Väänänen, A.K.P., Heponiemi, T., Salo, P., *et al.* (2013), "Chronic workplace stress and insufficient physical activity: a cohort study", *Occupational and Environmental Medicine*, BMJ Publishing Group Ltd, Vol. 70 No. 1, pp. 3–8, doi: 10.1136/oemed-2012-100808.

30. Le, H.T.K. and Poom, A. (2024), "Advancing environmental exposure and health impact assessment research with travel behaviour studies", *Transport Reviews*, Routledge, Vol. 44 No. 4, pp. 733–742, doi: 10.1080/01441647.2023.2297454.

31. Lee, C.-F., Lee, J.C. and Lee, A.C. (2013), "Analysis of Variance and Chi-Square Tests", in Lee, C.-F., Lee, J.C. and Lee, A.C. (Eds.), *Statistics for Business and Financial Economics*, Springer, New York, NY, pp. 543–612, doi: 10.1007/978-1-4614-5897-5_12.

32. Li, S., Liu, Z. and Zhao, H. (2020), "K-means Clustering Method Based on Kernel Density Estimation to Analysis Residents Travel Features: A Case Study of Chengdu", *Journal of Physics: Conference Series*, IOP Publishing, Vol. 1646 No. 1, p. 012018, doi: 10.1088/1742-6596/1646/1/012018.

33. Li, W., Zhang, Y., Chen, Y., Ding, L., Zhu, Y. and Chen, X. (Michael). (2024), "Multi-day activity pattern recognition based on semantic embeddings of activity chains", *Travel Behaviour and Society*, Vol. 34, p. 100682, doi: 10.1016/j.tbs.2023.100682.

34. Lin, H.-Z., Lo, H.-P. and Chen, X.-J. (2009), "Lifestyle classifications with and without activity-travel patterns", *Transportation Research Part A: Policy and Practice*, Vol. 43 No. 6, pp. 626–638, doi: 10.1016/j.tra.2009.04.002.

35. Ma, J., Liu, G., Kwan, M.-P. and Chai, Y. (2021), "Does real-time and perceived environmental exposure to air pollution and noise affect travel satisfaction? evidence from Beijing, China", *Travel Behaviour and Society*, Vol. 24, pp. 313–324, doi: 10.1016/j.tbs.2021.05.004.

36. Mahmoudi, J. (2022), "Health impacts of nonmotorized travel behavior and the built environment: Evidence from the 2017 National Household Travel Survey", *Journal of Transport & Health*, Elsevier, Vol. 26, p. 101404, doi: 10.1016/j.jth.2022.101404.

37. McCrorie, P., Mitchell, R., Macdonald, L., Jones, A., Coombes, E., Schipperijn, J. and Ellaway, A. (2020), "The relationship between living in urban and rural areas of Scotland and children's physical activity and sedentary levels: a country-wide cross-sectional analysis", *BMC Public Health*, Vol. 20 No. 1, p. 304, doi: 10.1186/s12889-020-8311-y.

38. National Bureau of Statistics of China. (2021), "The Seventh National Population Census Bulletin", available at: https://www.gov.cn/guoqing/2021-05/13/content_5606149.htm (accessed 5 June 2025).

39. Paggi, M.E., Jopp, D. and Hertzog, C. (2016), "The Importance of Leisure Activities in the Relationship between Physical Health and Well-Being in a Life Span Sample", *Gerontology*, Vol. 62 No. 4, pp. 450–458, doi: 10.1159/000444415.

40. Poom, A., Willberg, E. and Toivonen, T. (2021), "Environmental exposure during travel: A research review and suggestions forward", *Health & Place*, Vol. 70, p. 102584, doi: 10.1016/j.healthplace.2021.102584.

41. Posadzki, P., Pieper, D., Bajpai, R., Makaruk, H., Könsgen, N., Neuhaus, A.L. and Semwal, M. (2020), "Exercise/physical activity and health outcomes: an overview of Cochrane systematic reviews", *BMC Public Health*, Vol. 20 No. 1, p. 1724, doi: 10.1186/s12889-020-09855-3.

42. Rasheed, E.O., Khoshbakht, M. and Baird, G. (2021), "Time spent in the office and workers' productivity, comfort and health: A perception study", *Building and Environment*, Vol. 195, p. 107747, doi: 10.1016/j.buildenv.2021.107747.

43. Sharma, T. and Jain, S. (2023), "Nexus between built environment, travel behaviour and human health: An integrated framework to reinform transport system", *Journal of Cleaner Production*, Elsevier, Vol. 416, p. 137744, doi: 10.1016/j.jclepro.2023.137744.

44. SPARCK JONES, K. (1972), "A STATISTICAL INTERPRETATION OF TERM SPECIFICITY AND ITS APPLICATION IN RETRIEVAL", *Journal of Documentation*, MCB UP Ltd, Vol. 28 No. 1, pp. 11–21, doi: 10.1108/eb026526.

45. SPSSAU. (2025). SPSSAU [Version 25.0, Online Application Software]. Retrieved from https://www.spssau.com.

46. United Nations Department of Economic and Social Affairs, Population Division. (2019), "World Urbanization Prospects The 2018 Revision".

47. Vert, C., Gascon, M., Ranzani, O., Márquez, S., Triguero-Mas, M., Carrasco-Turigas, G., Arjona, L., *et al.* (2020), "Physical and mental health effects of repeated short walks in a blue space environment: A randomised crossover study", *Environmental Research*, Vol. 188, p. 109812, doi: 10.1016/j.envres.2020.109812.

48. Xu, J. (2001), "Self-rated health measurement scale, SRHMS", *Chinese Journal of Behavioral Medical Science*, No. 10 (special), pp. 25–7.

49. Xu J., Wang B., Hu M., Yang Y., Chen H. and Xie Y. (2000), "The develoment and evaluation of self-rated health measurement scale-prior test version", *Chinese Journal of Behavioral Medical Science*, No. 1.

50. Xu, W., Zeng, Z., Xu, Z., Li, X., Chen, X., Li, X., Xiao, R., *et al.* (2020), "Public health benefits of optimizing urban industrial land layout - The case of Changsha, China", *Environmental Pollution*, Vol. 263, p. 114388, doi: 10.1016/j.envpol.2020.114388.

51. Yan, Z. (2011), "Robust and Hierarchical Stop Discovery in Sparse and Diverse Trajectories".

52. Yang J., Li Z. and Liu Y. (2020), "Urban functional area identification method and its application integrating OSM road network data and POI data", *Geomatics World*, No. 6, pp. 1–6, 13.

53. Yang Y., Shangguan S. and Fang X. (2024), "The health effects of spatial accessibility of rural medical services from a fair perspective", *Chinese Rural Economy*, No. 11, pp. 102–124, doi: 10.20077/j.cnki.11-1262/f.2024.11.005.

54. Yang Z., Su J., Yang H. and Zhao Y. (2021), "Refinement of urban functional areas based on multi source data: a case study of Beijing", *Geographical Research*, No. 2, pp. 477–494.

55. Zhang J., Qiu P., Xu Z. and Du M. (2013), "A method to identify trip based on the mobile phone positioning data", *Journal of Wuhan University of Technology: Transportation Science & Engineering*, Vol. 37 No. 5, pp. 934–938.

56. Zhang, Y., Tzortzopoulos ,Patricia and and Kagioglou, M. (2019), "Healing built-environment effects on health outcomes: environment–occupant–health framework", *Building Research & Information*, Routledge, Vol. 47 No. 6, pp. 747–766, doi: 10.1080/09613218.2017.1411130.

57. ZHOU, Y. and YANG, C. (2018), "Anchors Identification in Trajectory Based on Temporospatial Clustering Algorithm", *Journal of Transportation Systems Engineering and Information Technology*, Vol. 18 No. 4, p. 88.

58. Zhu, J. and Fan, Y. (2018), "Daily travel behavior and emotional well-being: Effects of trip mode, duration, purpose, and companionship", *Transportation Research Part A: Policy and Practice*, Vol. 118, pp. 360–373, doi: 10.1016/j.tra.2018.09.019.

59. Zhu L. (2024), "Analysis and Research on the Change of Urban Functional Area Based on POI Data", *Geomatics AND Spatial Information Technology*, No. 6, pp. 186–189.