

An Analysis of Correlation between Personality and Visiting Place using Spearman's Rank Correlation Coefficient

Ha Yoon Song* and Seongjin Park

Department of Computer Engineering
Hongik University, Seoul, Korea

*Corresponding author: Ha Yoon Song
[E-mail: hayoon@hongik.ac.kr]

*Received February 12, 2019; revised October 16, 2019; accepted February 24, 2020;
published May 31, 2020*

Abstract

Recent advancements in mobile device technology have enabled real-time positioning so that mobile patterns of people and favorable locations can be identified and related researches have become plentiful. One of the fields of research is the relationship between the object properties and the favored location to visit. The object properties of a person include personality, which is a major property jobs, income, gender, and age. In this study, we analyzed the relationship between the human personality and the preference of the location to visit. We used Spearman's Rank correlation coefficient, one of the many methods that can be used to determine the correlation between two variables. Instead of using actual data values, Spearman's Rank correlation coefficient deals with the ranks of the two data sets. In our research, the personality and the location data sets are used. Our personality data is ranked in five ranks and the location data is ranked in 8 ranks. Spearman's Rank correlation coefficient showed better results compared to Pearson linear correlation coefficient and Kendall rank correlation coefficient. Using Spearman's correlation coefficient, the degree of the relationship between the personality and the location preference is found to be 43%.

Keywords: Personality-Location Relationship, Personality, Location Preference, Spearman's Rank Correlation

1. Introduction

Recent advancements in mobile device technology have enabled real-time positioning so that mobile patterns of people and favorable locations can be identified; furthermore, related researches have increased in number. One area of the research is the relationship between the object properties and the favored location to visit. The object properties of a person include personality, which is a major property, jobs, income, gender, and age. In this study, we analyze the relationship between the human personality and the pattern of the visited locations. Various methods exist that find the correlation between two variables, and we choose regression analysis. There are also various types of regression methods. In this study, Spearman's Rank correlation coefficient [1] is utilized with ranking data. Instead of using actual data values, Spearman's Rank correlation coefficient (SRCC) deals with the ranks of two sets. The coefficient can be obtained using equation 1.

$$p = \frac{\sum_i(x_i - \bar{x})(y_i - \bar{y})}{\sqrt{\sum_i(x_i - \bar{x})^2} \sqrt{\sum_i(y_i - \bar{y})^2}} \quad (1)$$

where x_i is the rank of the i th data of the random variable X , y_i is the rank of the i th data of the random variable Y , \bar{x} and \bar{y} denote the expectations of X and Y , respectively. In our research, the personality and the location data sets were used. The personality data was ranked for five different factors and the location data was ranked for 8 different categories. Using SRCC, the correlation between personality and location preference was determined. Other than SRCC, Pearson linear correlation coefficient (PLCC) and Kendall rank correlation coefficient (KRCC) are also frequently used and PLCC and KRCC are checked also with the same dataset. For each SRCC, PLCC and KRCC, the relationship between personality and location preference were calculated and SRCC showed best result among these three correlations, and thus we concentrated on SRCC throughout this research.

Numerous research results have arisen with the recent advancement of geo-positioning data acquisition methods with portable devices such as smartphones. However, there exist very few results regarding the quantitative relationship between personality and the preferred place of visit of a person. This may be due to the problem of personal information disclosure. Burbey showed results regarding the future location and the arrival time at the location of a person based on past mobility data [2]. P. T. Costa and R. R. McCrae Costa and McCrae presented a quantitative method in order to represent the human personality, the so-called Big Five Factors (BFF), which include openness, conscientiousness, extraversion, agreeableness, and neuroticism [3]. Song and Lee introduced stepwise regression to analyze the relationship between personality and favorable places, with personality as an independent variable and locations as the dependent variable [4]. Draper and Smith showed a helpful tool for examining the relationship between the BFF and the categorized locations [5]. Similar to the methods in Song and Lee [4], regression analysis was applied. Also, Song and Kang analyzed the

relationship between personality and favorable places, with personality as an independent variable and locations as the dependent variable, using various regression methods such as Poisson regression, ZINB regression, and Quantile regression [6]. Results Amichai-Hamburger and Vinitzky are also based on human personality and its relationship with social networks [7]. The effect of the human personality on Facebook usage is researched. Self-reports of participants are used to extract object criteria and measurements from Facebook data, and the result shows that a strong relationship exists between human personality and actions on Facebook. Using the results of the above research, many of the applications can be improved in various fields. For example, recommendation systems can utilize BFF to recommend travel destinations, restaurants for lunch, beer halls after dinner and so on, along with deducing the relationship for the favorable locations. In section 2, we will discuss data used in this paper, from the aspect of personality data and location data. We will discuss the process of data for the rank in section 3. In section 4, we will show the results of SRCC. In the final section 5, we will conclude this paper with future research directions.

2. Personality Data and Location Data

2.1. Personality Data

Personality can be numerically represented by the BFF model. The basic theory of the categories of personality can be found in the Five Factor Model (FFM)); a set of questionnaires called the Big Five Inventory (BFI) can be used to acquire the personality of a person. In this research, 30 volunteers provided their BFI results to find their BFF. BFF is composed of five different personality factors: openness (O), conscientiousness (C), extraversion (E), agreeableness (A), and neuroticism (N). These can be represented in numerical values from 0 to 5 and act as independent variables for regression analysis. Each person is anonymously represented. **Table 1** shows the BFF of 30 volunteers. For example, **Table 1** shows that Person3 has higher openness compared to other people, in a numerical manner. O, C, E, A, and N in **Table 1** are represented as follows.

- O : Openness
- C : Conscientiousness
- E : Extraversion
- A : Agreeableness
- N : Neuroticism

2.2. Location

Location data can be collected using portable devices. In this study, volunteers used to collect location data. Several applications provide functionalities to collect location and positioning data. For the collection of location data, Swarm Foursquare Labs, Inc. [8] is used, and for the collection of positioning data, Sports Tracker Sports Tracking Technologies [9] is used. Between these two sets of related data, the focus of this paper is on location data. **Table 2** shows part of the location data collected by Person12 using the Swarm application. **Table 2**

contains the location name and the count of visits, which are raw data collected by the Swarm application. Hundreds of locations can be found in the raw data. The classification of categories is from Korean Standard Statistical Classification provided by Statistics Korea (KOSTAT)[10]. In the beginning, the topmost ten categories according to the count of visits are applied in this research. **Table 3** shows categorized results based on raw data as shown in **Table 2** of Person12. For example, CGV Hongdae is the name of movie theater and thus categorized into Movie Theater among ten categories in **Table 3**. Ten categories are identified with category codes. For instance, the count of Institutions of High Education contains counts for the University Lab, the University Library, and the University Fields. However, after preprocessing of data set, only eight categories are applied in this research. Meaningless ordinary visits were filtered out in the stage of preprocessing.

3. Preprocessing of Data for Rank

We used two sorts of data: personality data in continuous values while location data in discrete values. For the discrete location data, PLCC is found not so suitable for this research while KRCC and SRCC are suitable coefficients for this research. It is found that SRCC is more suitable than KRCC as a result of this research.

SRCC requires ranked data. Location and personality data sets presented in Section 2 must be ranked a priori. In addition, several trimmings of the collected data are essential to find meaningful results. For example, a very small number of count of visits collected by a specific person can be excluded, as it will not affect the rank. As well, other preprocessing is also done for too many visits to ordinary locations. For example, category of "Institutions of High Education" is removed from data set since it is an ordinary visit with too many counts when students visit university. With a similar reason, the category of "Ground Transportation Service" is also removed. Of course, the related data is also removed for these removed categories as well as personal data which only contains these removed data sets. In other words, we removed meaningless data from our data set.

Tables 4 and 5 show the resulting ranks based on the data from **Tables 1 and 3**, respectively. For the ranks of the BFF, five ranks are adequate, and for the eight location categories, 8 ranks are adequate. **Table 4** shows the ranked values of the BFF for each volunteer. The highest factor is assigned a value of 1, meaning the highest rank, whereas the lowest factor is assigned the value 5, meaning the lowest rank. For example, the highest valued factors of Person1 is conscientiousness and therefore has a ranking value of 1, while the lowest valued factor, neuroticism, has a ranking value of 5. It is possible that two factors have the same value. In such cases, an adjustment of rank values is required. For example, Person5 has the same value of 4.00 for O and E, as shown in **Table 1**, which are the first and second ranks. In this case, the average of the two ranks, 1.5 is assigned as the rank value as shown in **Table 4**. In addition, **Table 5** shows the ranked values of eight locations. The location with the highest visiting count has rank 1, while the location with the lowest visiting count has rank 8.

Table 1. Personality Data

Volunteer/ Factor	O	C	E	A	N
Person1	3.300	3.889	3.250	3.667	2.625
Person2	2.700	3.222	3.250	2.667	2.750
Person3	4.333	3.125	2.250	3.200	2.889
Person4	4.200	4.333	3.500	3.556	2.625
Person5	4.000	3.667	4.000	3.889	2.750
Person6	3.500	3.778	3.375	3.222	3.000
Person7	2.200	3.444	3.000	3.111	2.625
Person8	2.600	2.778	3.375	3.111	2.625
Person9	3.400	3.222	3.375	3.333	3.125
Person10	3.100	3.667	3.375	3.222	3.500
Person11	3.400	3.556	3.625	2.889	2.500
Person12	3.800	4.000	3.125	3.778	2.250
Person13	3.000	3.444	3.250	2.444	3.000
Person14	2.400	3.444	3.500	3.333	3.125
Person15	3.600	3.333	2.875	3.111	3.500
Person16	2.000	2.889	3.875	2.778	2.875
Person17	3.400	3.556	3.375	2.778	2.875
Person18	3.000	3.333	3.250	2.556	2.875
Person19	3.800	3.556	3.750	2.889	2.750
Person20	3.100	3.444	3.500	2.444	2.875
Person21	3.800	3.667	3.250	3.000	3.125
Person22	3.400	3.000	3.000	2.889	2.500
Person23	3.100	3.000	3.250	3.333	3.250
Person24	2.800	3.778	3.750	3.333	2.250
Person25	3.100	3.444	3.000	3.333	2.875
Person26	3.200	3.778	3.375	3.444	3.375
Person27	2.800	3.222	3.500	2.889	2.500
Person28	3.800	3.778	3.500	3.222	3.000
Person29	3.100	3.111	3.500	2.667	2.875
Person30	3.100	3.111	3.500	2.667	2.875

Table 2. Location Data

Location	Count
University Lab	132
CGV Hongdae	24
Hongik Hospital	10
Seoul Express Bus Terminal	7
Beer Warehouse	5
Deungchon Balling House	3

Table 3. Macro Location Data

Location	Category Code	Count
Foreign Institutions	4700	247
Large Retail Business	4711	379
Ground Transportation Service	5291	1297
Restaurant	5611	786
Bar	5621	320
Beverage Store	5622	270
Movie Theater	5914	24
Institutions of High Education	8530	320
Hospital	8610	495
Museums and Historical Sites	9022	63

4. Results of Spearman's Rank Correlation Coefficient

SRCC for personality data and location data from the rank shown in Section 3 needs to be found. Equation 2 shows the calculation of the SRCC.

$$r_s = \rho_{rg_x, rg_y} = \frac{cov(rg_x, rg_y)}{\sigma_{rg_x} \sigma_{rg_y}} \quad (2)$$

ρ denotes the general Pearson correlation and is only applied to rank variables. The covariance of the rank variables is $cov(rg_x, rg_y)$, the standard deviations of the rank variables are σ_{rg_x} and σ_{rg_y} respectively, and X and Y are random variables of ranks. A simpler form of the SRCC can be found in equation 3.

$$r_s = 1 - \frac{6 \sum d_i^2}{n(n^2 - 1)} \quad (3)$$

where n is count of variables and d denotes the difference between each pair of variables. Equation 3 is used in this paper. **Table 6** shows the results of the SRCC for 10 volunteers out of the 30 volunteers. Similarly, **Table 7** shows the results of SRCC of locations for 10 volunteers out of the 30 volunteers.

The legends in **Table 7** can be translated as follows:

- A : Foreign Institute
- B : Large Retail Business
- C : Restaurant
- D : Bar
- E : Beverage Store
- F : Movie Theater
- G : Hospital
- H : Museums and Historical Sites

Table 4. Ranks of Personality Data

Volunteer/ Factor	O	C	E	A	N
Person1	3	1	4	2	5
Person2	4	2	1	5	3
Person3	1	3	5	2	4
Person4	2	1	4	3	5
Person5	1.5	4	1.5	3	5
Person6	2	1	3	4	5
Person7	5	1	3	2	4
Person8	5	3	1	2	4
Person9	1	4	2	3	5
Person10	5	1	3	4	2
Person11	3	2	1	4	5
Person12	2	1	4	3	5
Person13	3.5	1	2	5	3.5
Person14	5	2	1	3	4
perosn15	1	3	5	4	2
Person16	5	2	1	4	3
Person17	2	1	3	5	4
Person18	3	1	2	5	4
Person19	1	3	2	4	5
Person20	3	2	1	5	4
Person21	1	2	3	5	4
Person22	1	2.5	2.5	4	5
Person23	4	5	2.5	1	2.5
Person24	4	1	2	3	5
Person25	3	1	4	2	5
Person26	5	1	3.5	2	3.5
Person27	4	2	1	3	5
Person28	1	2	3	4	5
Person29	3	2	1	5	4
Person30	3	2	1	5	4

Fig. 1 and **Fig. 2** are visualization of data from **Tables 6 and 7** in a form of bar chart. These bars stand for ranks, therefore the shortest bar shows the highest rank. For example, **Table. 1** and **Fig. 1** show BFF of Person1 respectively where neuroticism (N) has the smallest value. Thus neuroticism (N) of Person1 has the fifth rank among BFFs and represented as the longest

bar.

For **Tables 6 and 7**, the lowest row shows the SRCC for personality data and location data, respectively. Because we have 30 volunteers, 435(30C2) cases can be found for the correlation coefficient of personality data and location data. Owing to the limitation of space, **Tables 6 and 7** show values only for 10 people. It is known that the value of SRCC is in the range of $[-1, 1]$. For example, **Table 6** shows the correlation coefficient of Person 2 as -0.8. It means that the personalities of Person2 and Person3 are highly negatively related, (i.e. highly opposed).

Table 5. Ranks of Location Data

Location	Category Code	Rank
Foreign Institutions (A)	4700	6
Large Retail Business (B)	4711	3
Restaurant (C)	5611	1
Bar (D)	5621	4
Beverage Store (E)	5622	5
Movie Theater (F)	5914	8
Hospital (G)	8610	2
Museums and Historical Sites (H)	9022	7

Table 6. Spearman Rank Correlation Coefficient: Personality Aspect

	Person 1	Person 2	Person 3	Person 4	Person 5	Person 6	Person 7	Person 8	Person 9	Person 10
O	3	4	1	2	1.5	2	5	5	1	5
C	1	2	3	1	4	1	1	3	4	1
E	4	1	5	4	1.5	3	3	1	2	3
A	2	5	2	3	3	4	2	2	3	4
N	5	3	4	5	5	5	4	4	5	2
Corr. Coeff-icient	-0.2	-0.8	0.6	0.21	0.36	0.3	0.6	0	-0.8	0.1

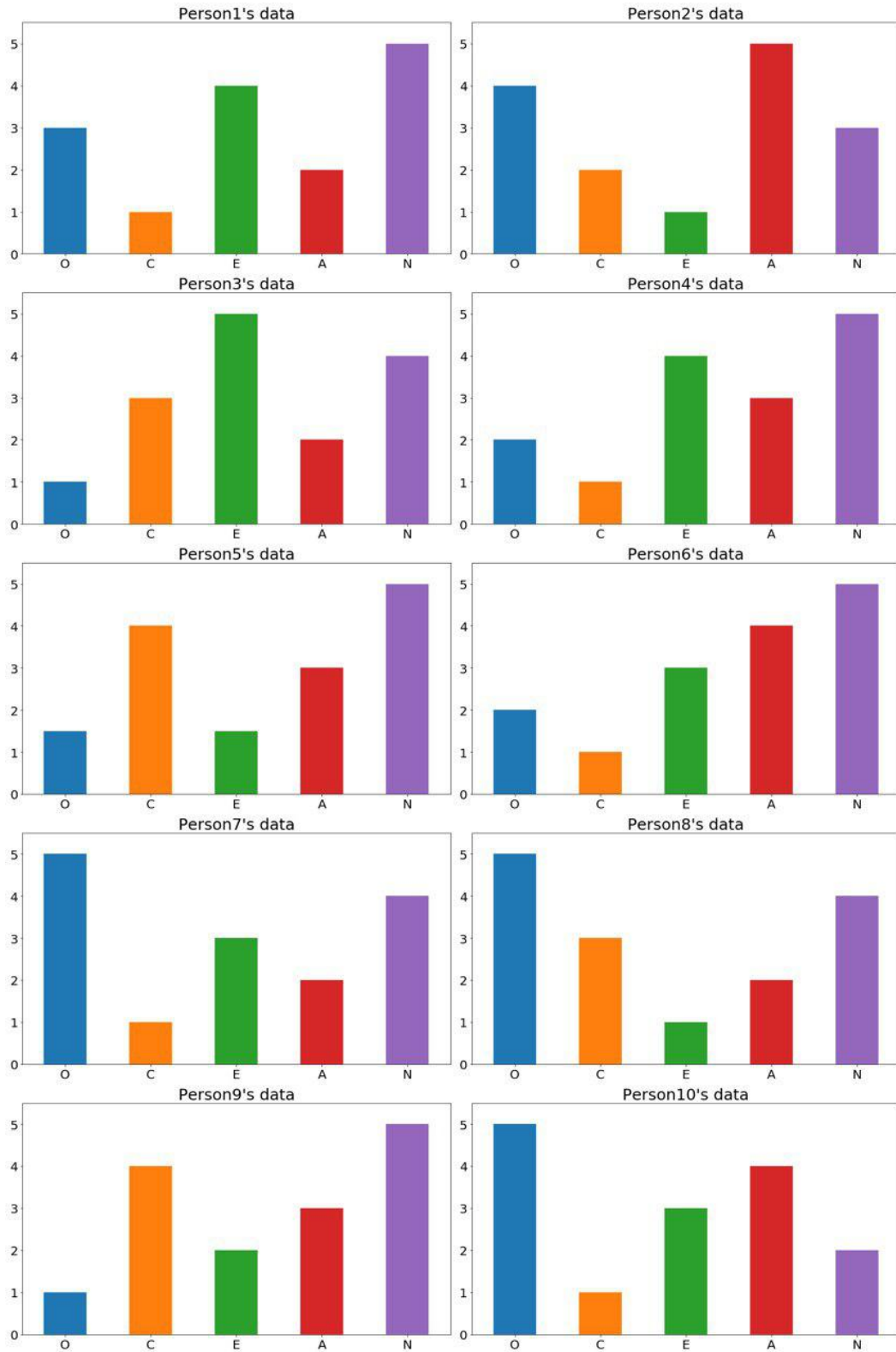


Fig. 1. Visualize of Table 6, Ranked Personality Data.

Table 8 and **Fig. 3** shows 30 < location coefficient, personality coefficient > pairs out of the possible 435(30C2) combinations. One < location coefficient, personality coefficient > pair is for one comparison of two persons. The third pair, < 0.79, 0.6 > shows a difference of 0.19 and can be deduced to be highly related. In these two cases, we can conclude that personality and location preference is proven to be highly related by SRCC. However, there is also a contradictory case. For example, the second pair < 0.49, -0.8 > shows that the opposite personality pattern can show related location preference. In other words, there is a case in which an opposite personality can lead to a similar locational preference. For all the 435 cases, graphical representation can be found in **Fig. 4**. **Fig. 4** has a personality correlation coefficient on the Y-axis and location correlation coefficient on the X-axis. In order to compare directly between SRCC, KRCC and PRCC results, we used linear regression. A linear regression between the two variables found a regression function of $y = 0.4399x + 0.05552$. The line in **Fig. 4** has the value of the slope as 0.4399 which implies that the degree of the relationship between the location correlation coefficient and the personality correlation coefficient is 43%.

Fig. 5 shows scatter and results of the comparison. For eight categories of location data, results of PLCC, KRCC, and SRCC are shown. As discussed in section 3, use of PLCC is inadequate for our location data in discrete values as shown in Fig. 4.

Rank correlations such as SRCC and KRCC show slopes of 0.4399 and 0.3911 respectively, while PLCC show slope of -0.03156 which is almost meaningless.

Table 7. Spearman Rank Correlation Coefficient: Locational Aspect

	Person 1	Person 2	Person 3	Person 4	Person 5	Person 6	Person 7	Person 8	Person 9	Person 10
A	6.5	5.5	2	1	3	5	6.5	6.5	7	7.5
B	3	1	4	5	5	4	3	3	3	1
C	1	3	1	2	1	1	1	2	1	3
D	4	4	6	4	4	3	4	8	4	5.5
E	2	2	3	3	2	2	2	1	2	2
F	6.5	5.5	7	6	6	6.5	6.5	5	5	4
G	8	8	5	8	7	6.5	5	6.5	8	5.5
H	5	7	8	7	8	8	8	4	6	7.5
Corr. Coeff-icient	0.83	0.49	0.79	0.9	0.92	0.92	0.54	0.71	0.76	0.04

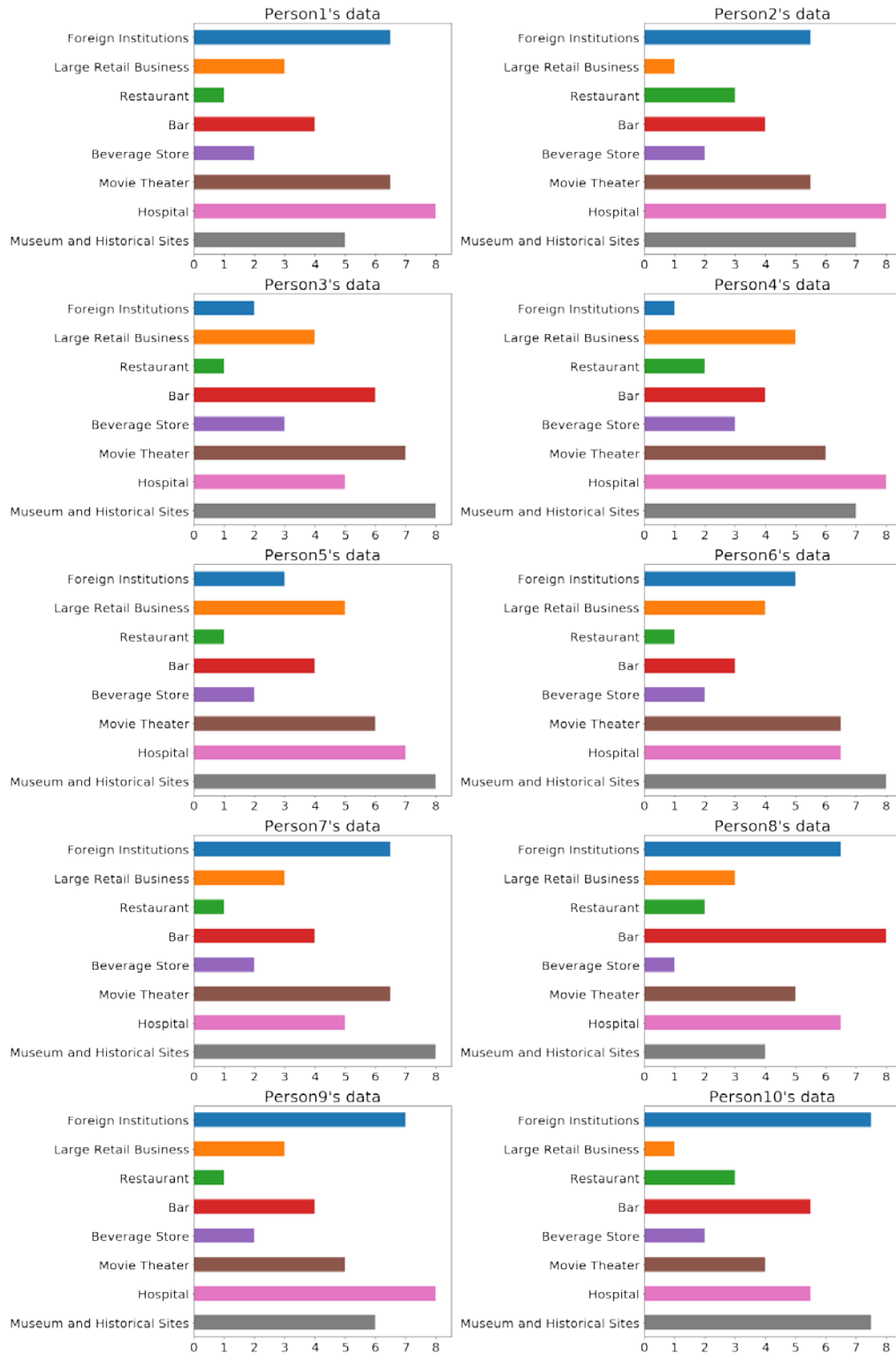


Fig. 2. Visualize of Table 7, Ranked Location Data.

Table 8. Correlation Coefficient between Personality and Location

Location coefficient	Personality coefficient	Location coefficient	Personality coefficient	Location coefficient	Personality coefficient
0.83	-0.2	0.24	0.4	0.2	0.87
0.49	-0.8	0.12	0.36	0.03	-0.61
0.79	0.6	0.81	0.56	0.22	-0.36
0.9	0.21	0.52	-0.9	0.59	0.7
0.92	0.36	0.76	-0.7	0.62	0.67
0.92	0.3	0.82	0.2	0.5	0.41
0.54	0.6	0.69	0.9	0.52	0.3
0.71	0	0.68	0.5	0.31	0.5
0.76	-0.8	0.52	0.6	0.79	1
0.04	0.1	0.02	0.6	0.47	0.5

5. Conclusion

Using Spearman's Rank correlation coefficient (SRCC), the degree of the relationship between location preference and personality has been studied in this paper. The location preference and personality factors were ranked and then SRCC were found. In the case of personality data, five ranks were applied, whereas ten ranks were applied for location data. Using the ranked data, the degree of the relationship between personality and location preference was observed to be 43%. As opposed to other researches Song and Lee [4]; Draper and Smith [5]; Song and Kang [6] which deal with the direct relationship between personality and specific locations or location categories, our research shows the degree of the relationship. The count of the location data seems to be sufficient to apply regression analysis, while we are doubtful of the count of the personality data, which is provided by 30 volunteers. A more precise result could be deduced with a greater volume of personality data. In addition, the number of ranks can be adjusted to find a more meaningful relationship. For example, if we have more detailed categories of restaurants such as Chinese restaurants, Indian restaurants, and so on, we could have more meaningful results while current categorization of a restaurant is not so detail and may be regarded as ordinary visits. In the future, we plan to conduct a more in depth research with these two considerations and to find more appropriate regression methods.

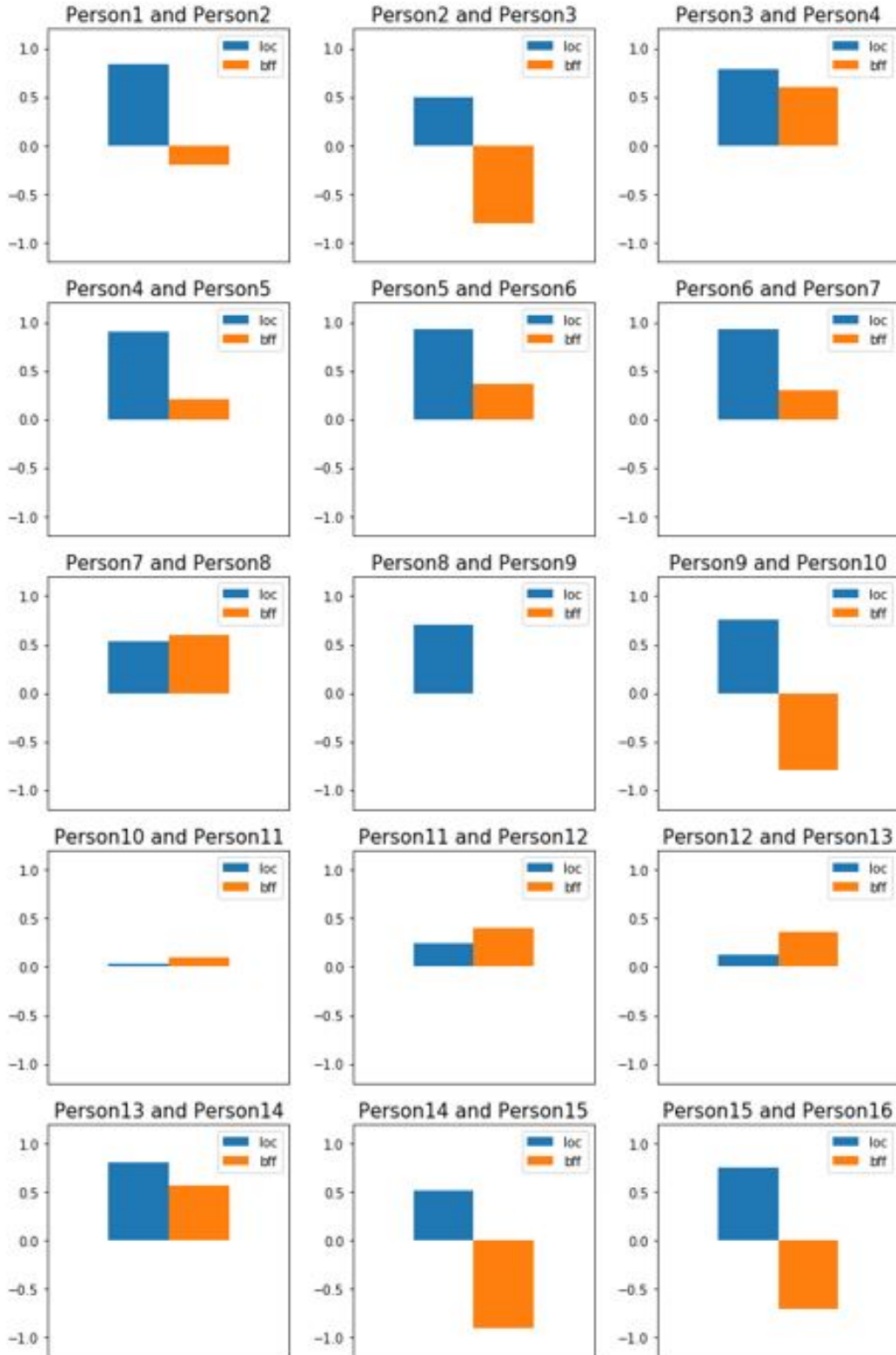


Fig. 3-a. Visualize Table 8. Correlation Coefficient between Personality and Location

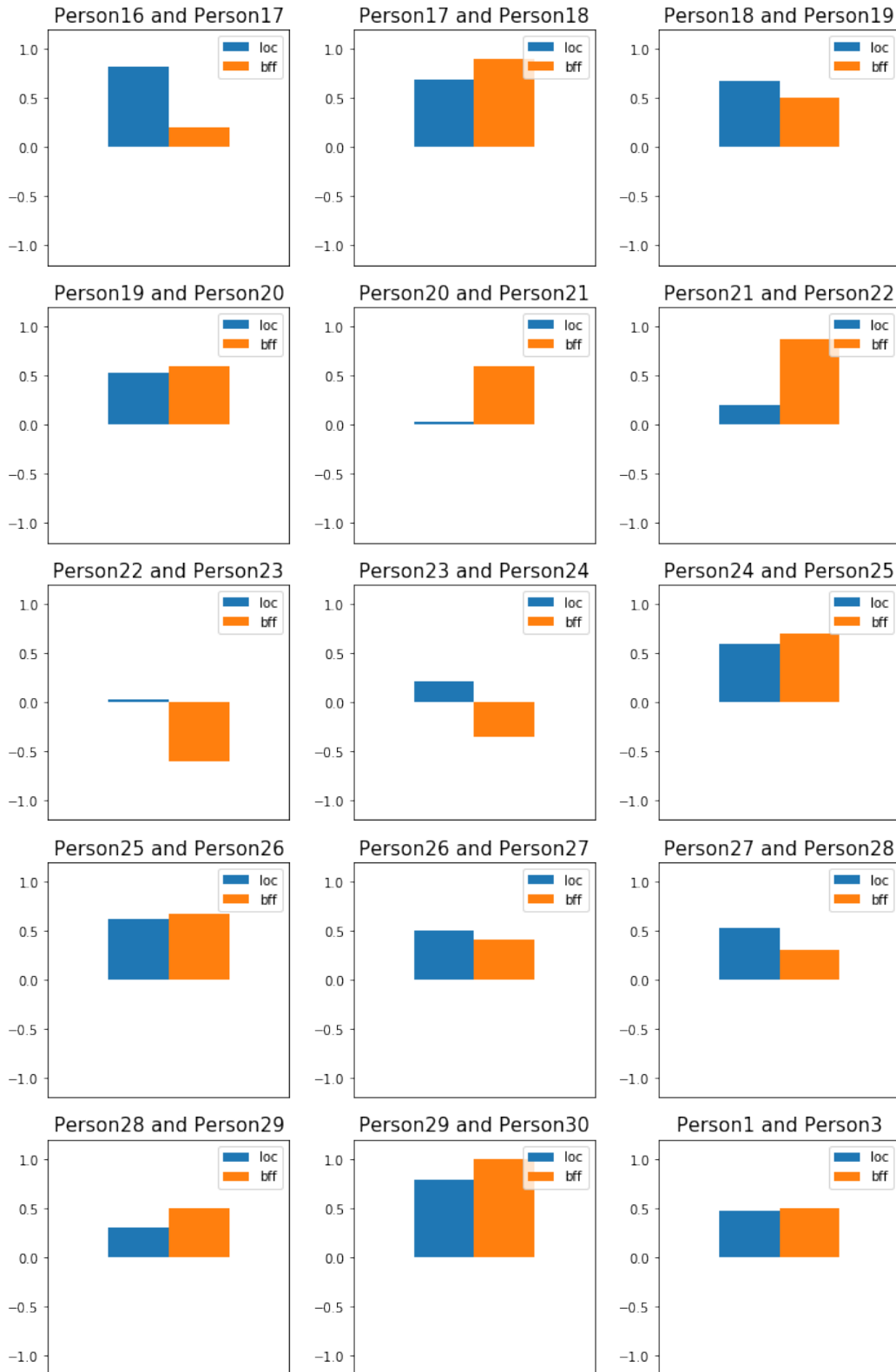


Fig. 3-b. Visualize Table 8. Correlation Coefficient between Personality and Location

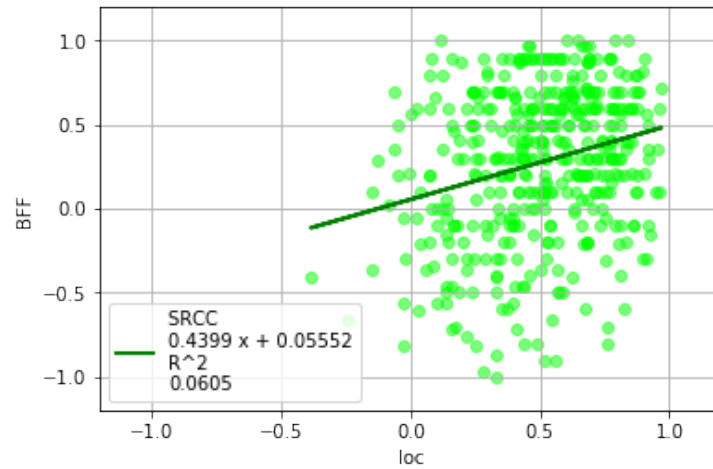


Fig. 4. Correlation between Personality and Location with SRCC

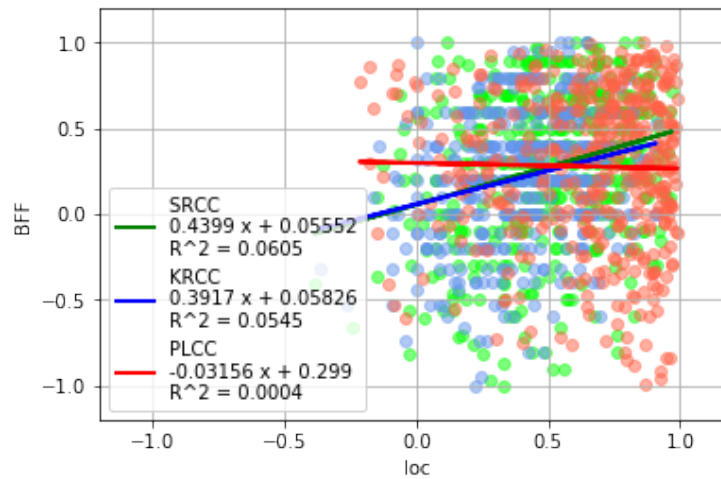


Fig. 5. Correlation between Personality and Location

Acknowledgement

This work was supported by the National Research Foundation of Korea (NRF) grant funded by the 306 Korea government (MEST) (NRF-2019R1F1A1056123).

References

- [1] Leann Myers and Maria J Sirois, "Spearman correlation coefficients, differences between," *Wiley StatsRef: Statistics Reference Online*, 2006. [Article \(CrossRef Link\)](#).
- [2] Ingrid Burbey, "Predicting future locations and arrival times of individuals," *PhD thesis*, 2011.
- [3] Paul T Costa and Robert R McCrae, "Four ways five factors are basic," *Personality and individual differences*, 13(6), 653–665, 1992. [Article \(CrossRef Link\)](#).
- [4] Ha Yoon Song and Eun Byul Lee, "An analysis of the relationship between human personality and favored location," in *Proc. of AFIN 2015*, pp. 12, 2015. [Article \(CrossRef Link\)](#).

- [5] Norman R Draper and Harry Smith, "Applied regression analysis," *John Wiley & Sons*, 2014.
- [6] Ha Yoon Song and Hwa Baek Kang. Analysis of relationship between personality and favorite places with poisson regression analysis," in *Proc. of The 2017 International Conference Applied Mathematics, Computational Science and Systems Engineering Athens, Greece*, vol. 16, 2018. [Article \(CrossRef Link\)](#).
- [7] Yair Amichai-Hamburger and Gideon Vinitzky, "Social network use and personality," *Computers in human behavior*, 26(6), 1289–1295, 2010. [Article \(CrossRef Link\)](#).
- [8] Foursquare Labs, Inc. Swarm app. <https://www.swarmapp.com>, 2019.
- [9] Sports Tracking Technologies. Sportstracker. <http://www.sports-tracker.com>, 2019.
- [10] Korean Standard Industrial Classification. https://kssc.kostat.go.kr:8443/ksscNew_web/ekssc/main/main.do#



Ha Yoon Song received his B.S. degree in Computer Science and Statistics in 1991 and received his M.S. degree in Computer Science in 1993, both from Seoul National University, Seoul, Korea. He received Ph.D. degree in Computer Science from University of California at Los Angeles, USA in 2001. From 2001 he has worked at Department of Computer Engineering, Hongik University, Seoul, Korea and is now a full professor. In his sabbatical year 2009, he worked at Institute of Computer Technology, Vienna University of Technology, Austria as a visiting scholar and in the year 2018, he worked at Free University Berlin as a visiting scholar. Prof. Song's research interests are in the areas of data science, mobile computing, application of machine learning, location based service and human mobility modeling.



Seongjin Park received his B.S. degree in Computer Engineering in 2019 and is expected to receive his M.S. degree in Computer Engineering in 2021, respectively, both from Hongik University, Seoul, Korea. Mr. Park's research interests are in the areas of deep learning, AutoML, location based service and human personality modeling.