

Math 5545 Project 2 (Due: March 16, 2021)

Principal Component Analysis of Kansas City Neighborhood Data

Group # 1: During-COVID – Warm Season

* Ford, William Andrew waftx2@mail.umkc.edu
** Tran, Thao Phuong ttcww@mail.umkc.edu
Johnson, Reece LaVaughn rlj6pb@mail.umkc.edu (Analysis of Low-income data)
Whetsell, Torsten P torsten.whetsell@mail.umkc.edu (Analysis of Moderate -income data)
Salas, Jacob R jrsty8@mail.umkc.edu (Analysis of High-income data)

Group # 2: Pre-COVID – Cold Season

* Vaughn, Braeden bv2my@mail.umkc.edu
** Stack, Caston A cas8y5@mail.umkc.edu
Terry, Harrison Edward Lee het9t5@mail.umkc.edu (Analysis of Low-income data)
Schaeffer, Alex as9nb@mail.umkc.edu (Analysis of High -income data)

Group # 3 Pre-COVID – Warm Season

* Lim, Celine Shwu Ling cs19r3@mail.umkc.edu
** Reesman, Grace Ellen gekhx5@mail.umkc.edu
Rosenblatt, Jennifer jrosenblatt@mail.umkc.edu (Analysis of Low-income data)
Thomas, Micheal mctwcc@mail.umkc.edu (Analysis of Moderate-income data)
Aljofei, Maha Mashan mmawv6@umkc.edu (Analysis of High-income data)

* **team leader.** The team leader is responsible for organizing the zoom meetings, distributing the work between the team members coordinating with the project manager and making sure that the project is getting finished before the deadline. If you do not receive an email from your team leader by the end of this week, please let me know by Monday morning.

** **project manager.** the project manager is responsible for the quality of the project, making sure all calculations are correct, coordinating with the team leader and submitting all required work.

Collegiality and Group Work

The groups in this class are meant to imitate real-world research groups. Each group should regularly meet via zoom. Each group member should (1) maintain a friendly environment for the entire group; (2) facilitate collaboration and problem solving; (3) provide a vision of the main objectives and ensure discussions lead to conclusions and decisions; (4) motivate and inspire other group members; (5) contribute to the group by sharing his/her knowledge, expertise, and viewpoints; (6) participate in all meetings and discussions; (7) have productive suggestions.

Principal Component Analysis of Kansas City Neighborhood Data: Measuring Health Index of Each Kansas City Neighborhood

Note: This is not the title of your project: You need to choose a title appropriate to the results that you obtain from analysis of KC Neighborhood data

Background

In order to study social health inequalities, contextual (or ecologic) data may constitute an appropriate alternative to individual socioeconomic characteristics. Indices can be used to summarize the multiple dimensions of the neighborhood socioeconomic status. This work proposes a statistical procedure to create a neighborhood socioeconomic index [1].

Objectives:

A) Reduce the dimension of the variables

With a large number of socioeconomic and demographic variables, the dispersion matrix may be too large to study and interpret properly. There would be too many pairwise correlations between the variables to consider. Graphical displays may also not be particularly helpful when the data set is very large. With 18 variables, for example, there will be more than 200 three-dimensional scatterplots.

B) Rank Kansas City neighborhoods based on new variables

To interpret the data in a more meaningful form, it is necessary to reduce the number of variables to a few, interpretable linear combinations of the data. Each linear combination will correspond to a principal component. We would like to interpret the data based on the new variables obtained after reducing the dimension.

Data & Variables:

There are two sets of data: (1) new data: excel file PCA-KC, and (2) Old data from project #1.

- (1) New data: The Excel file PCA-KC data will be used for the principal component analysis (PCA). As listed below, the data set of 200 KCMO neighborhoods has 17 variables.

NBH_ID NBH_NAME

1. total crime
2. violent crime
3. property crime
4. Median income
5. Median home value
6. Population with BS or More: Total population age 25+ years with a bachelor's degree or higher
7. White alone
8. Black or African American
9. alone Asian
10. alone Hispanic or Latino

11. Total Vacant
12. Total Renter Occupied
13. Number of Households
14. Walk Score
15. Transit Score
16. Bike Score
17. Population

(2) Old data: You can obtain this data from canvas project number one.

COVID-19 interval of 3/1/2020 – 9/1/2020 warm season – **Assigned to Group 1**

Pre-COVID-19 interval of 9/1/2019 – 3/1/2020 cold season – **Assigned to Group 2**

Pre-COVID-19 interval of 3/1/2019 – 9/1/2019 warm season – **Assigned to Group 3**

Key Facts about KC MO

1. Kansas City has a Livability Score of 58/100, which is considered poor
2. Kansas City crime rates are 69% higher than the Missouri average
3. Cost of living in Kansas City is 4% higher than the Missouri average
4. Kansas City real estate prices are 4% lower than the Missouri average
5. Rental prices in Kansas City are 9% higher than the Missouri average

<https://www.areavibes.com/kansas+city-mo/?ll=39.09973+-94.57857>

Methodology for calculating Walk Score, transit score and Bike Score,

90–100 Walker's Paradise; Daily errands do not require a car.

70–89 Very Walkable; Most errands can be accomplished on foot.

50–69 Somewhat Walkable; Some errands can be accomplished on foot.

25–49 Car-Dependent; Most errands require a car.

0–24 Car-Dependent; Almost all errands require a car.

90–100 Rider's Paradise; World-class public transportation

70–89 Excellent Transit; Transit is convenient for most trips

50–69 Good Transit; Many nearby public transportation options

25–49 Some Transit; A few nearby public transportation options

0–24 Minimal Transit; It is possible to get on a bus

90–100 Biker's Paradise; Daily errands can be accomplished on a bike

70–89 Very Bikeable; Biking is convenient for most trips

50–69 Bikeable; Some bike infrastructure

0–49 Somewhat Bikeable; Minimal bike infrastructure

Score Methodology: <https://www.walkscore.com/methodology.shtml>

Source of Data for columns Walk Score, transit score, Bike Score, and population

: https://www.walkscore.com/MO/Kansas_City

Methodology for calculating the crime scores: in this study the crime scores were calculated according to the national average 4 total crime, violent crime and property crime. A number in the first three Columns of the data represents percentage above or below the national average.

Crime data was manually obtained using the website <https://www.areavibes.com/kansas+city-mo/crime/> the numbers indicate percentage above or below national average

Additional notes/considerations:

The crime data is not pure; there are some inconsistencies in the data.

There were 15 neighborhoods that we were not able to find the crime data. we set the crime values to the average Kansas City Mo crime values of 112% higher than national average for total crime, 277% higher than national average in violent crime, 83% higher than national average in property crime

Procedure:

Part A) Apply PCA to new data

1. Watch this video first:
<https://umkc.hosted.panopto.com/Panopto/Pages/Viewer.aspx?id=9c9cd26a-cc1e-4140-807c-acdf013910db>
2. Replicate the steps Lesson 11 and Example 11-1 provided in the following website:
<https://online.stat.psu.edu/stat505/lesson/11>
3. **Important Note:** Make sure to include the math behind the PCA in your slides.
4. **Software:** You can use SPSS available in the UMKC remote lab. For specific instruction please visit: <https://statistics.laerd.com/spss-tutorials/principal-components-analysis-pca-using-spss-statistics.php>
5. **It is highly recommended that each group member individually tries the software and that you compare your results**

Part B) Use the old data to add new variables to the new data

Using the old data, fill out the last five columns of the Excel file PCA-KC Data-Assignment. You must compute average day to close for each one of the following categories for each of 200 neighborhoods:

1. Animals / Pets
2. City Facilities
3. Lights / Signals
4. Mowing / Weeds
5. Trash / Recycling

Part C) Repeat part A for the new dataset generated in Part B.

References:

1. Neighborhood PCA analysis: <https://equityhealthj.biomedcentral.com/articles/10.1186/1475-9276-12-21>
2. Wright, R., M. Heilweil, P. Pelletier and K. Dickinson (1974). The Impact of Street Lighting on Crime. Ann Arbor, Mich.: Institute for Social Research, University of Michigan. (#197)
3. Sherman, L., and D. Rogan (1995). "Effects of Gun Seizures on Gun Violence: 'Hot Spots' Patrol in Kansas City." Justice Quarterly 12(4):673-693. (#166)
4. Crime data was manually obtained using the website <https://www.areavibes.com/kansas+city-mo/crime/> the numbers indicate percentage above or below national average
5. Additional crime data were found from : <https://www.weichert.com/search/community/neighborhood.aspx?hood=73306>
6. See for example the crime data for Columbus Park Kansas City, MO <https://www.weichert.com/search/community/neighborhood.aspx?hood=6176>
7. KC Neighborhoods: <https://www.areavibes.com/kansas+city-mo/neighborhoods/>
8. KC Crime data: <https://www.areavibes.com/kansas+city-mo/crime/>
9. KC Stat: <https://www.kcmo.gov/home/showdocument?id=288>
10. Dangerous: KC Neighborhoods Crime: <https://www.areavibes.com/kansas+city-mo/most-dangerous-neighborhoods/>
11. Safe: KC Neighborhoods Crime: <https://www.areavibes.com/kansas+city-mo/safest-neighborhoods/>
12. Key Facts about KC MO <https://www.areavibes.com/kansas+city-mo/?ll=39.09973+-94.57857>
13. Crime comparison: <https://www.areavibes.com/crime-comparison/>
14. Neighborhood component analysis; note that this is different than PCA https://en.wikipedia.org/wiki/Neighbourhood_components_analysis