CS 584-04: Machine Learning

Fall 2018 Assignment 4

# Question 1 (60 points)

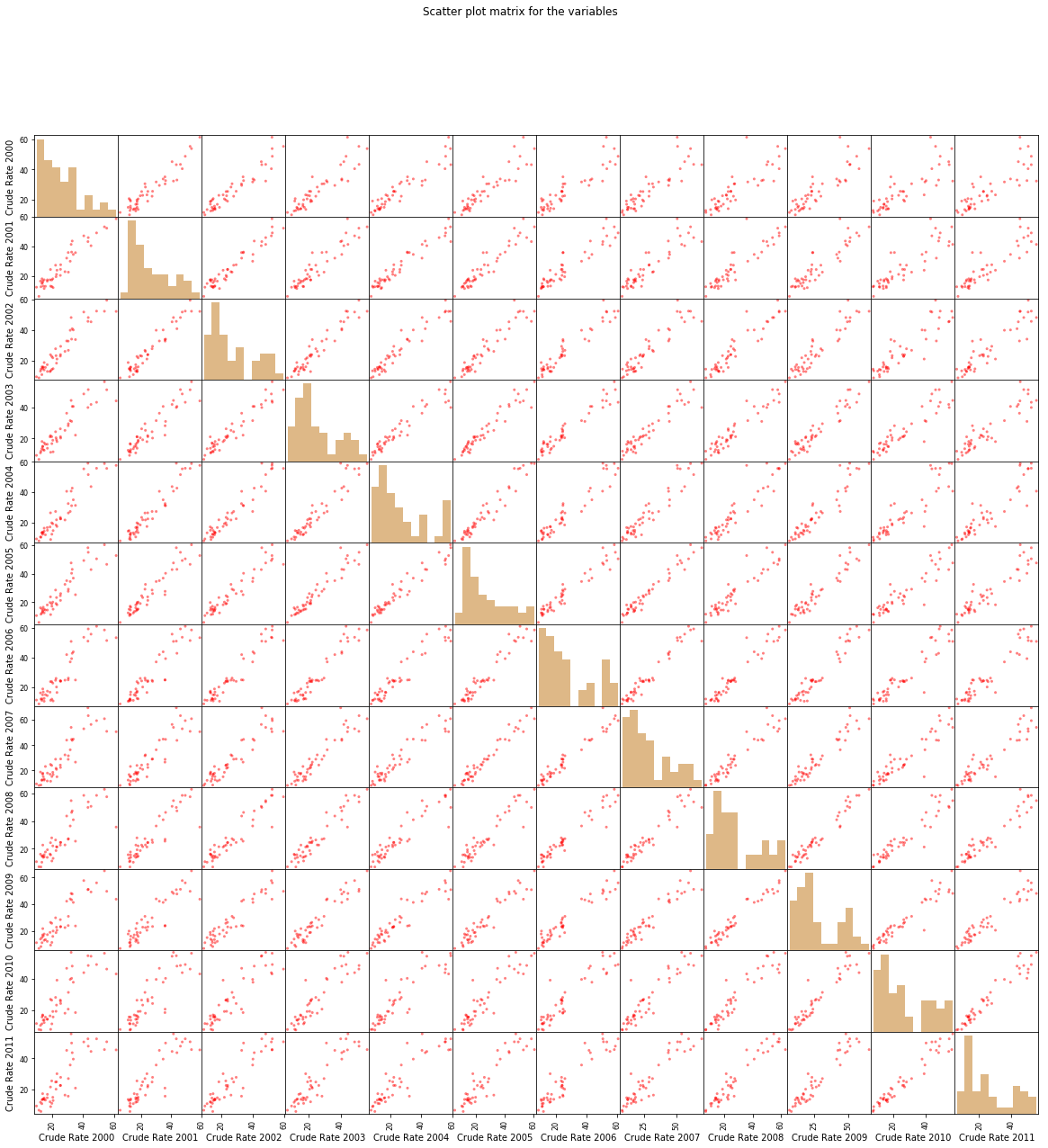
1. (5 points). How many observations and variables did you use in your Principal Component analysis?

Shape of data: [46 rows x 12 columns]

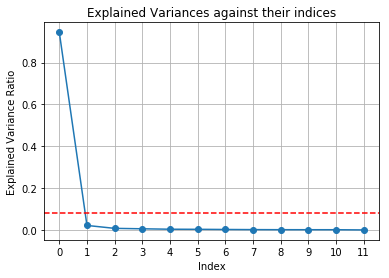
46 observations and 12 variables are used

1. (5 points). Generate the scatter plot matrix for the variables. Put the histograms on the diagonal.

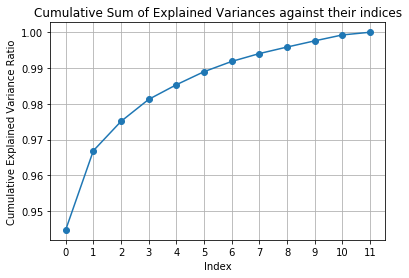
----------------------------------------------Scroll to next page---------------------------------------------------------



1. (5 points). Plot the Explained Variances against their indices. Add a horizontal reference line whose value is the reciprocal of the number of variables. Label the axes and add grid lines to the axes.



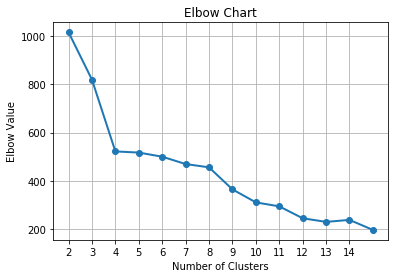
1. (5 points). Plot the Cumulative Sum of the Explained Variances against their indices. Label the axes and add grid lines to the axes.

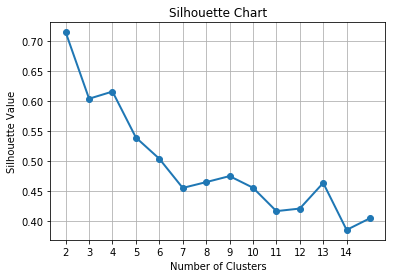


1. (5 points). What percentage of the total variance is explained by the first two principal components?

96.690181%

1. (5 points). Plot the Elbow and the Silhouette charts against the number of clusters.





1. (5 points). What is the number of clusters that you will choose based on the charts in f)?

4

1. (5 points). How many communities are in each cluster?

Cluster ID count

0 18

2 14

1. 9
2. 5
3. (5 points). List the names of the communities in each cluster.

Cluster 0:

0 60601, 60602, 60603, 60604, 60605 & 60611 Downtown

1 60606, 60607 & 60661 West Loop

4 60610 & 60654 Near North Side

6 60613 Lake View

7 60614 Lincoln Park

11 60618 Avondale

18 60625 Albany Park

22 60630 Jefferson Park

23 60631 Edison Park

24 60632 Archer Heights

25 60634 Dunning

31 60641 Portage Park

34 60645 West Ridge

35 60646 Edgebrook

41 60655 Mount Greenwood

42 60656 Norwood Park

43 60657 Belmont Harbor

44 60659 North Park

Cluster 1:

5 60612 Near West Side

10 60617 South Chicago

27 60637 Woodlawn

32 60643 Beverly

38 60651 West Humboldt Park

Cluster 2:

2 60608 Lower West Side

3 60609 New City

8 60615 Hyde Park

9 60616 Chinatown

15 60622 & 60642 West Town

16 60623 South Lawndale

19 60626 Rogers Park

21 60629 West Lawn

28 60638 Garfield Ridge

29 60639 Belmont Gardens

30 60640 Edgewater

36 60647 Bucktown

39 60652 Ashburn

45 60660 Edgewater Glen

Cluster 3:

12 60619 Chatham

13 60620 Auburn Gresham

14 60621 Englewood

17 60624 West Garfield Park

20 60628 Roseland

26 60636 West Englewood

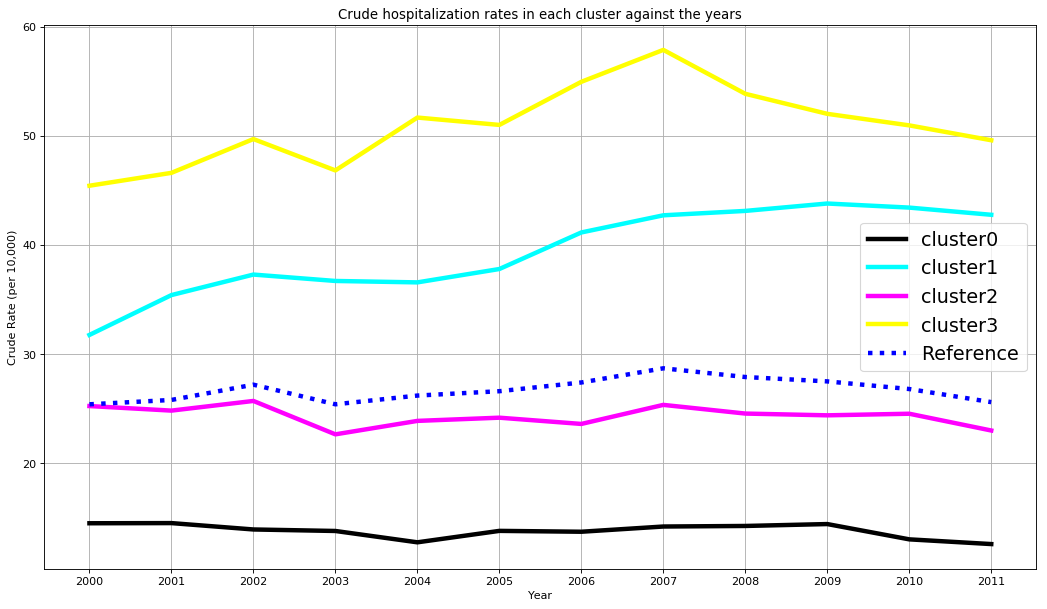
33 60644 Austin

37 60649 South Shore

40 60653 Kenwood

1. (10 points). Plot the crude hospitalization rates in each cluster against the years. You also plot the Chicago’s annual crude hospitalization rates (in the table below) against the years as the reference curve.

|  |  |  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| Year | 2000 | 2001 | 2002 | 2003 | 2004 | 2005 | 2006 | 2007 | 2008 | 2009 | 2010 | 2011 |
| Rate | 25.4 | 25.8 | 27.2 | 25.4 | 26.2 | 26.6 | 27.4 | 28.7 | 27.9 | 27.5 | 26.8 | 25.6 |



1. (5 points) Based on the graph in j), what will you conclude about the trend of crude hospitalization rate in each cluster relative to the Chicago’s rates?

Chicago’s rates haven’t increased nor decreased drastically albeit with a few fluctuations. All clusters see a rise in crude rate in the year 2007. Cluster 3 neighborhoods have the highest crude hospitalization rates followed by cluster 1, both of which are higher than the reference. Cluster 2 and cluster 0 have lower rates than the reference.

# Question 2 (40 points)

1. (5 points) What are the Class Probabilities?

Class A = 0 probability 0.21599581510081187

Class A = 1 probability 0.64046244338586

Class A = 2 probability 0.1435417415133281

1. (5 points) When group\_size = 1, homeowner = 0, and married\_couple = 0, what are the predicted probabilities Pr(A = 0), Pr(A = 1), and Pr(A = 2)?

Pr(A = 0) is **0.2697221048499267**

Pr(A = 1) is **0.5801329399401**

Pr(A = 2) is **0.15014495520997334**

1. (5 points) When group\_size = 2, homeowner = 1, and married\_couple = 1, what are the predicted probabilities Pr(A = 0), Pr(A = 1), and Pr(A = 2)?

Pr(A = 0) is **0.13827430813010222**

Pr(A = 1) is **0.7259545618746951**

Pr(A = 2) is **0.1357711299952027**

1. (5 points) When group\_size = 3, homeowner = 1, and married\_couple = 1, what are the predicted probabilities Pr(A = 0), Pr(A = 1), and Pr(A = 2)?

Pr(A = 0) is **0.19436967601795363**

Pr(A = 1) is **0.6404093169902023**

Pr(A = 2) is **0.16522100699184422**

1. (5 points) When group\_size = 4, homeowner = 0, and married\_couple = 0, what are the predicted probabilities Pr(A = 0), Pr(A = 1), and Pr(A = 2)?

Pr(A = 0) is **0.37549062583572096**

Pr(A = 1) is **0.4878096506845897**

Pr(A = 2) is **0.1366997234796893**

1. (10 points) What are the values of the predictors group\_size, homeowner, and married\_couple such that Prob(A = 1) attains its maximum?

Group Size = 2

Homeowner = 1

Married Couple = 1

1. (5 points) For the values of group\_size, homeowner, and married\_couple, what are the predicted probabilities Pr(A = 0), Pr(A = 1), and Pr(A = 2)?

The below data (predicted probabilities) for each of the values of the variables is in the format:

(Pr(A = 0), Pr(A = 1), Pr(A = 2))

group\_size= 1 homeowner= 0 married\_couple= 0

(0.2697221048499267, 0.5801329399401, 0.15014495520997334)

group\_size= 1 homeowner= 0 married\_couple= 1

(0.2327894062536048, 0.6142181021428793, 0.152992491603516)

group\_size= 1 homeowner= 1 married\_couple= 0

(0.1940380809567132, 0.6696585761357657, 0.13630334290752105)

group\_size= 1 homeowner= 1 married\_couple= 1

(0.16493516175037906, 0.6982776295037999, 0.13678720874582112)

group\_size= 2 homeowner= 0 married\_couple= 0

(0.23114351470080471, 0.6165180049554043, 0.15233848034379088)

group\_size= 2 homeowner= 0 married\_couple= 1

(0.19801576122393472, 0.6479063332453118, 0.15407790553075348)

group\_size= 2 homeowner= 1 married\_couple= 0

(0.1636276819009433, 0.700287393833978, 0.1360849242650787)

group\_size= 2 homeowner= 1 married\_couple= 1

(0.13827430813010222, 0.7259545618746951, 0.1357711299952027)

group\_size= 3 homeowner= 0 married\_couple= 0

(0.3082195887875182, 0.5159236953280671, 0.17585671588441468)

group\_size= 3 homeowner= 0 married\_couple= 1

(0.268311240184029, 0.5509504199790198, 0.18073833983695112)

group\_size= 3 homeowner= 1 married\_couple= 0

(0.2269720106552963, 0.6096113185495722, 0.16341667079513142)

group\_size= 3 homeowner= 1 married\_couple= 1

(0.19436967601795363, 0.6404093169902023, 0.16522100699184422)

group\_size= 4 homeowner= 0 married\_couple= 0

(0.37549062583572096, 0.4878096506845897, 0.1366997234796893)

group\_size= 4 homeowner= 0 married\_couple= 1

(0.33074366964210417, 0.5270978482121912, 0.1421584821457047)

group\_size= 4 homeowner= 1 married\_couple= 0

(0.28217290202526524, 0.5881960077903524, 0.12963109018438237)

group\_size= 4 homeowner= 1 married\_couple= 1

(0.24393054429629368, 0.6237655229384779, 0.13230393276522845)