

# LBP Patients Clustering

## 2017 IFCS Data Challenge

Fengmei Liu, Sucharu Gupta  
Math 285 Project



**SAN JOSÉ STATE**  
UNIVERSITY

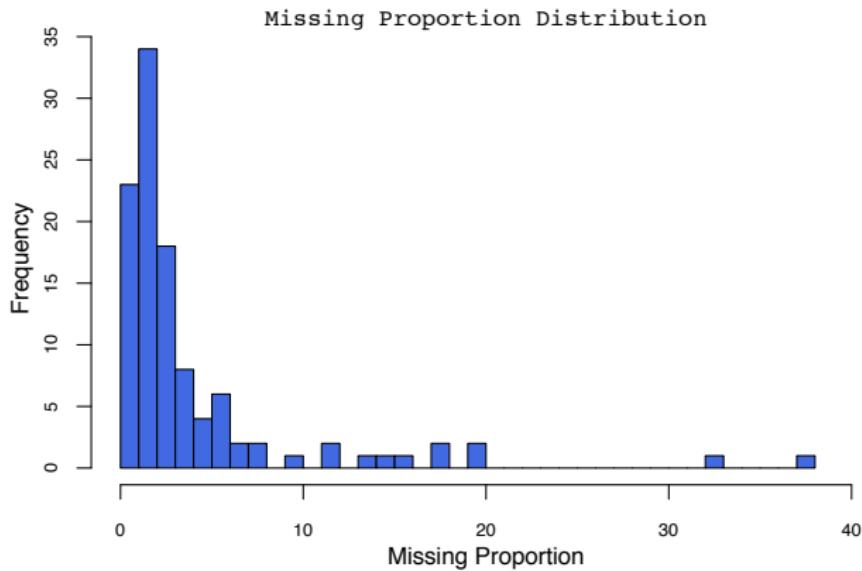
# Data Detailing

- Data

data\_challenge2.xlsx dimensions: 928 \* 122

variables\_data\_challenge2.xlsx dimensions: 122 \* 14

- Missing Values



# Data Summarizing

- Domain

1	Activity	23
2	Contextual factors	16
3	Pain	14
4	Participation	8
5	Physical impairment	24
6	Psychological	27

- Data Types

1	Continuous	8
2	Dichotomous	64
3	Multistate nominal	9
4	Ordinal	30
5	Trichotomous	1

# Data Glitches and Feature Selection

- Imputed missing with valid values for appropriate variables  
Students/Pensioners/Unemployed/: 67 to 75  
LBP dominate or not: 86 to 89
- Change data to related types
- Summary scores not included: 85, 122
- Missing values over 20% not included: 91, 98
- If any level dominate one variable, using 85% cutoff line  
exclude: 35 57 69 93 94 95 100 109 111 112 113 114 115
- Imputation  
R Mice package, random forest method

# Clustering - Multiple Correspondence Analysis (MCA) + Reduced Kmeans (RKM)

Process:

- Transform all continuous variables to categorical
  - Age
  - Bhoej0/Height
  - Vasl0/LBP Intensity
  - Okon0/Able to decrease pain
  - Obeh0/Treatment not essential
  - Htl0/Self-rated general health
  - bmi/BMI
- MCA clustering on 6 domains
- Overall RKM using the reduced output of 6 domains(12 new columns)

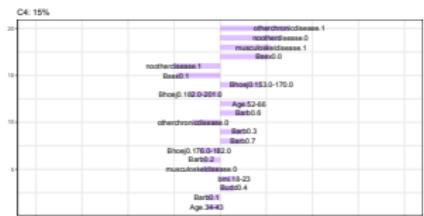
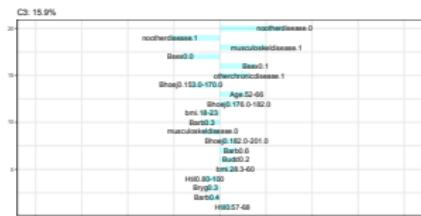
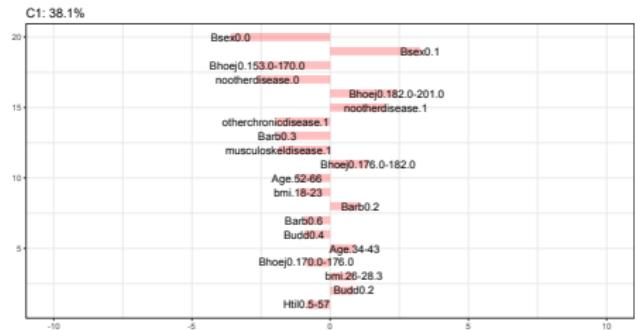
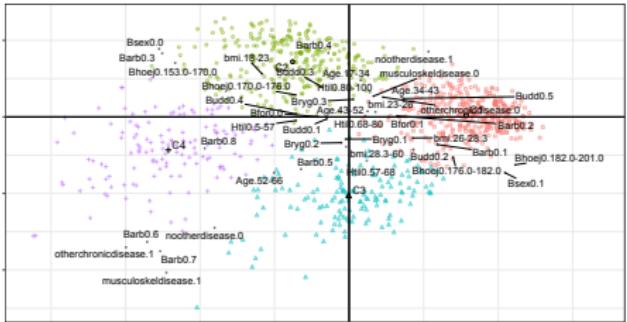
# STEP1 MCA Results - Contextual Factor Domain

X1: + male, age34-43, tall, no chronic disease, full-time work, low education

- female, old, short, has musculoskel/other chronic disease, retired, high education

X2: + female, young, tall, no other chronic disease, work, high education.

- male, old(52-66), average height, has musculoskel/other chronic disease,(low education)



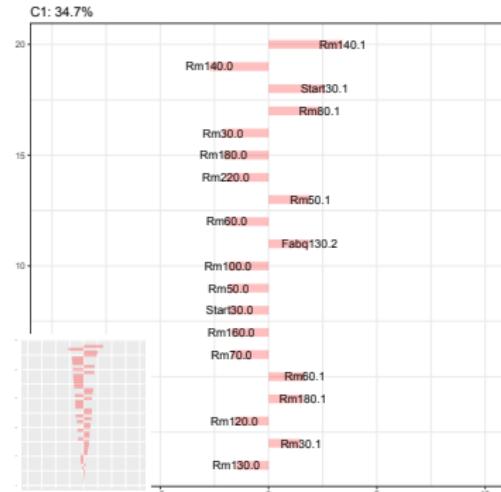
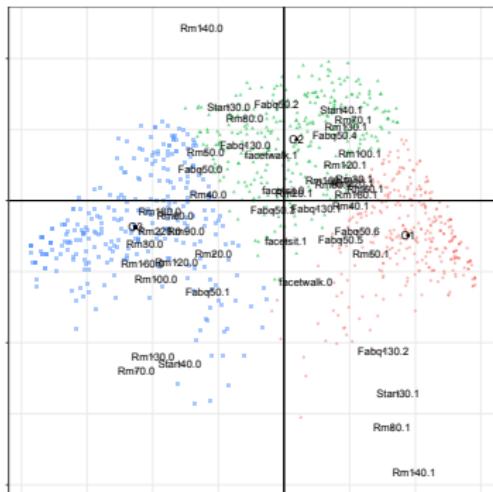
# STEP1 MCA Results - Activity Domain

X3: + home activity slowly, self dress slowly

- home activity normal, self dress normal

X4: + walk short distance, stand for short time, self-dress normal

- walk normal, stand normal, self-dress slowly



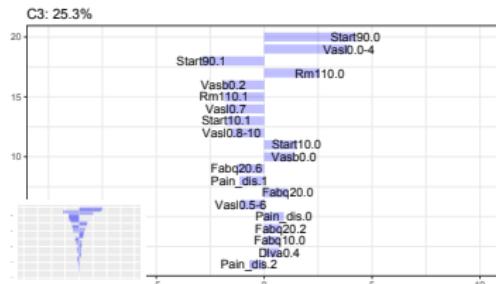
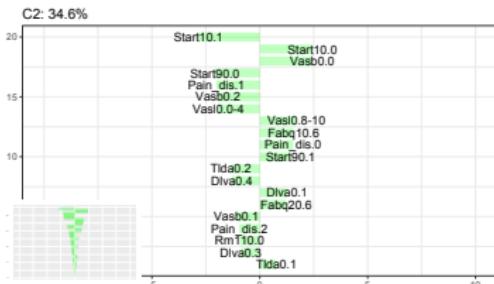
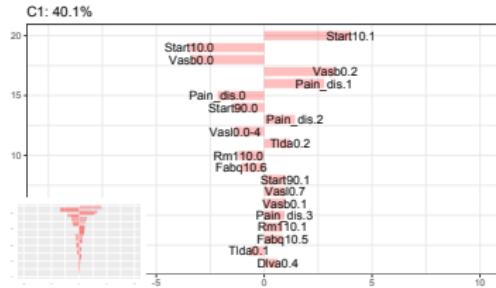
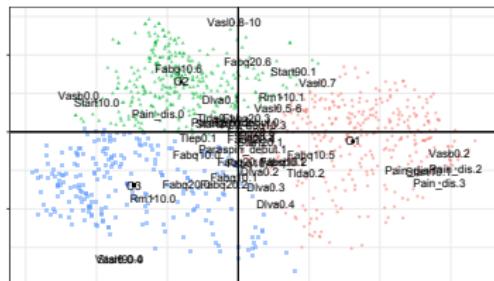
# STEP1 MCA Results - Pain Domain

X5: + pain has spred dwon to legs, leg pain is intense

- pain has not spred down to legs, no leg pain

X6: + LBP pain intensity is high, this pain episode last short

- LBP pain intensity is low, this pain episode last long



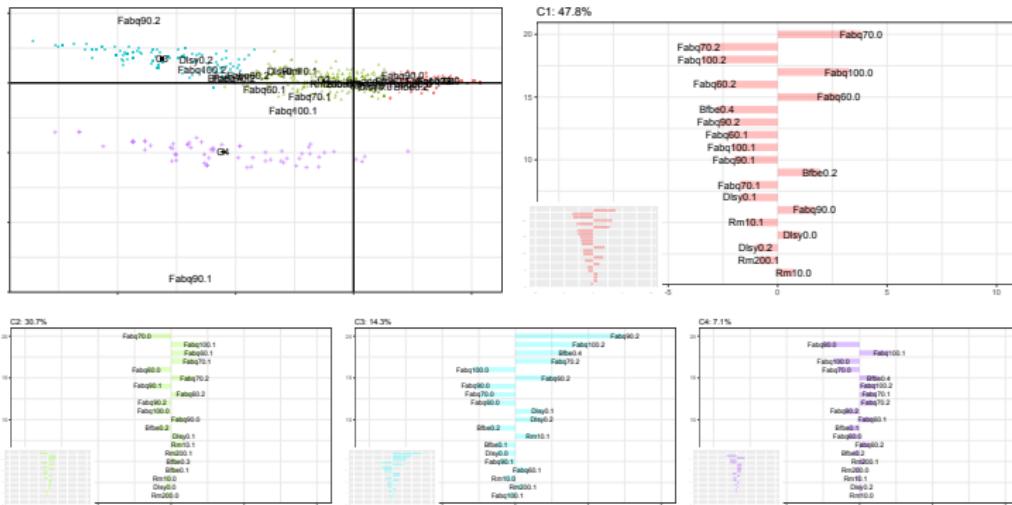
# STEP1 MCA Results - Participation Domain

X7: + work not make pain worse, phisical work load sitting/walking

- work make pain worse, heavey phisical work

X8: + work is heavey, more sick leave and stay home time.

- unsure if work is heavy, less sick leave and stay home time



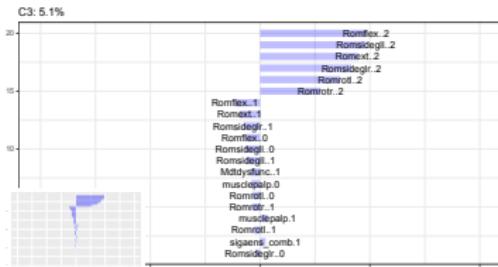
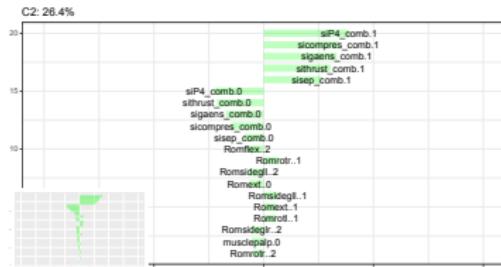
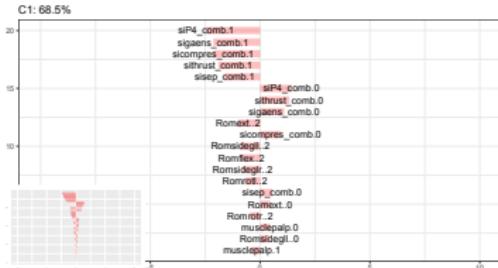
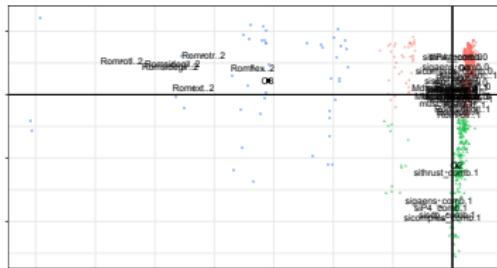
# STEP1 MCA Results - Physical Impairment Domain

X9: + no pain on AROM

- leg pain on AROM test

X10: + negative on SI-joint tests, no pain on palpation

- positive on SI-joint test, back pain on AROM test

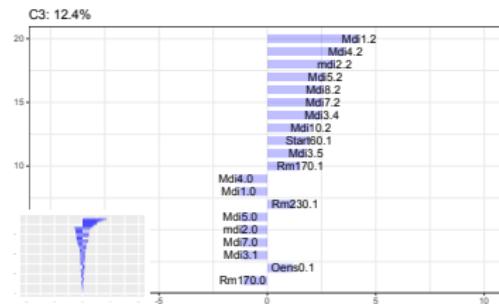
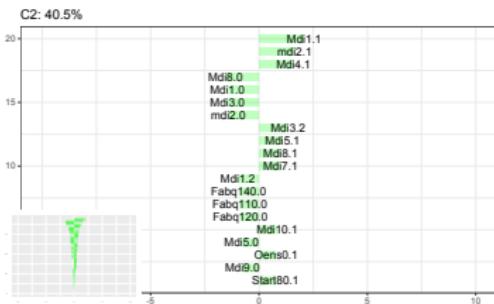
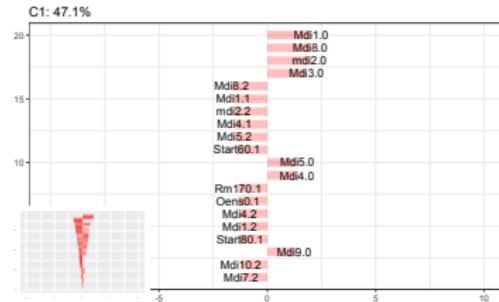
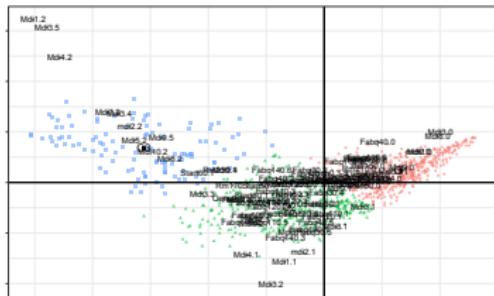


# STEP1 MCA Results - Psychological Domain

X11 + good mode and feel energy, good conscience, good sleep

- bad mode and feel less energy, bad conscience, bad sleep

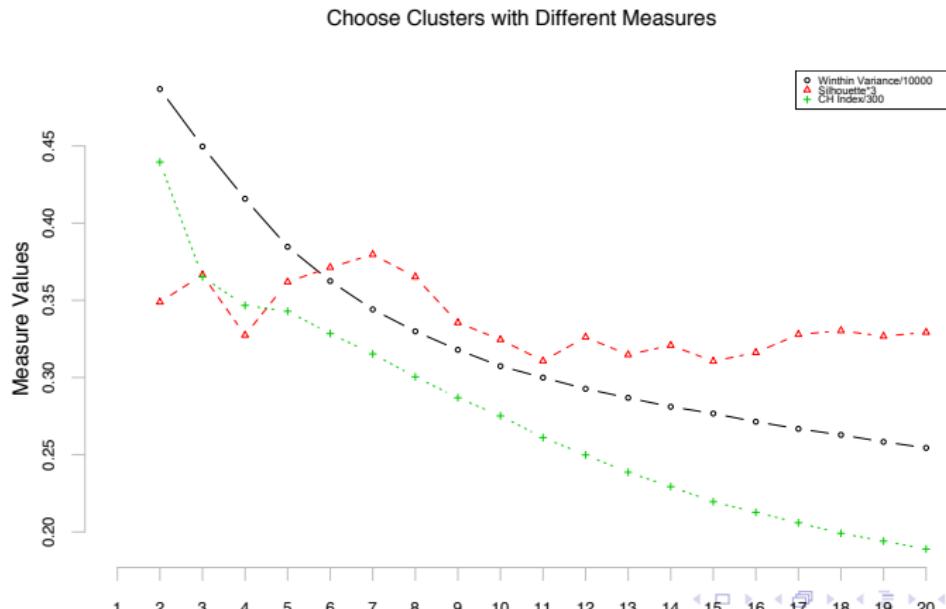
X12 + not lose interest in daily activities, psychologically believe should do activity - lose interest in daily activities, psycho believe should not do activities



# STEP2 RKM

Tried various methods, chose Reduced Kmeans to implement the step2 clustering. Chose the number of clusters based on Within Variance, Silhouette width and CH index.

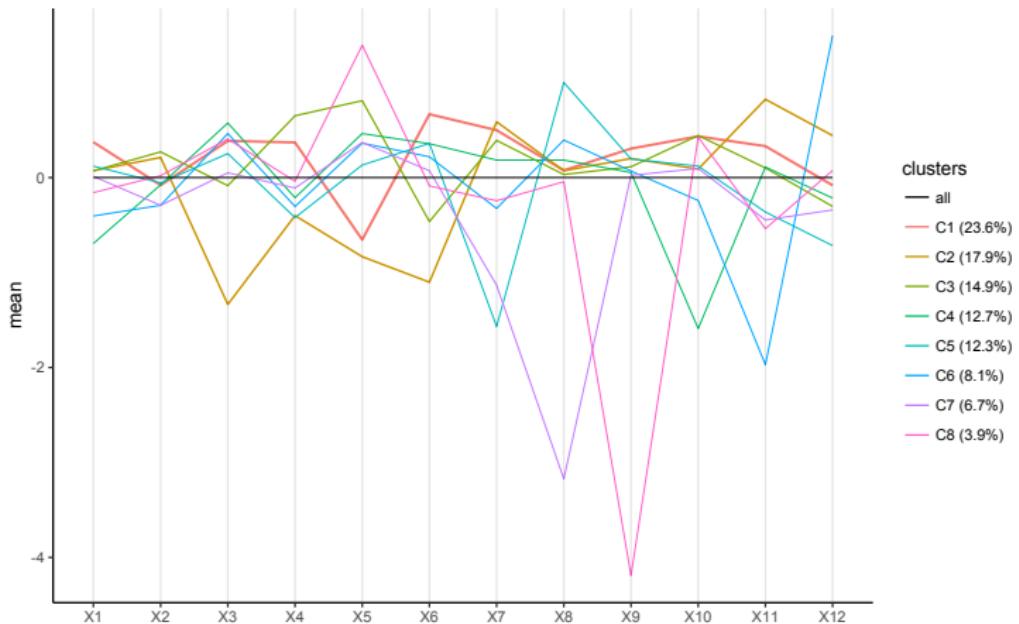
Groups with range 5-8 are OK, we finalized with 8 groups.



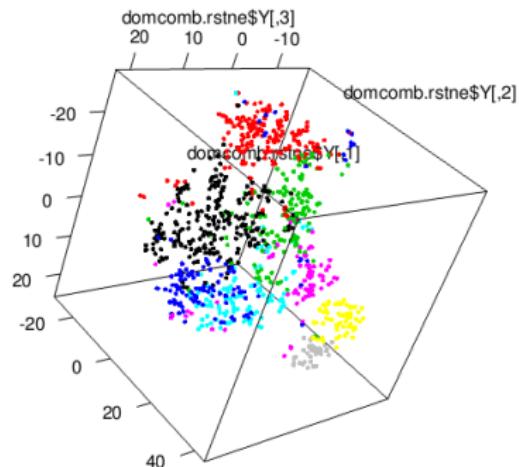
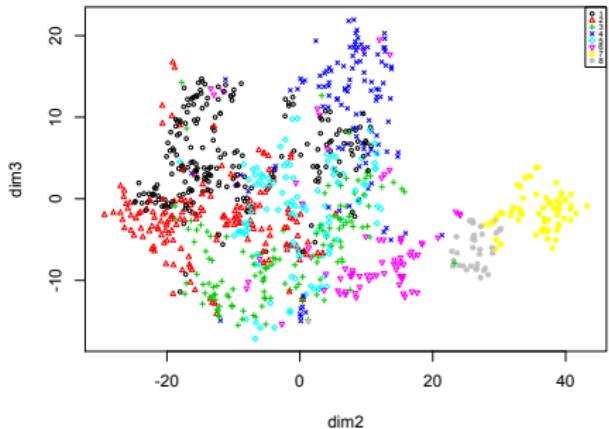
# STEP2 RKM - Result

E.g. C1:

male, age34-43, no chronic disease, full-time work; walk short distance, stand for short time, self-dress normal; pain has not spread down to legs; LBP pain intensity is high; This pain episode last short; Work not make pain worse, physical work load sitting/walking; no pain on AROM test; negative on SI-joint tests, no pain on palpation; good mode and feel energy, good sleep;

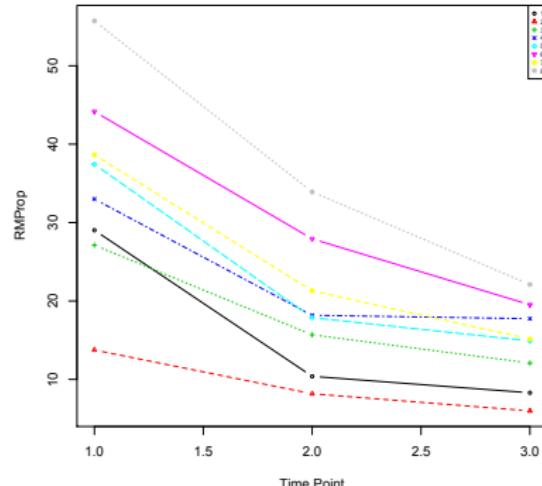
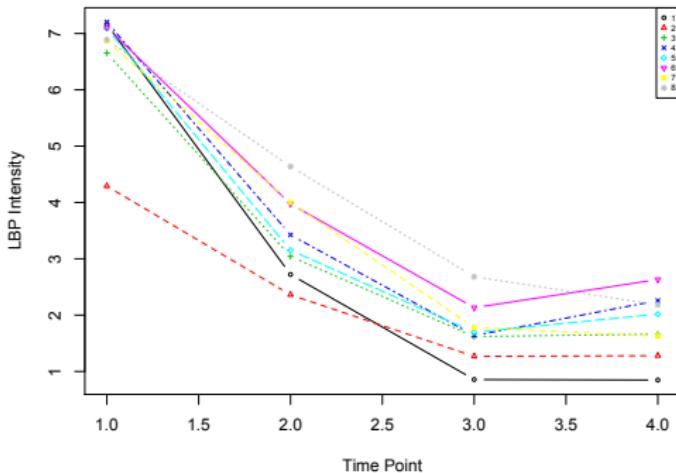


# STEP2 RKM Result



Visualize the clustering result using dimension reduction method TSNE. Clusters are shown and well separated.

# Clustering Results Interpretation



E.g. Cluster 1, 2 are lowest, in our results, cluster1 is middle-ages men who have intense LBP with a short duration, so it healed quickly and dropped down fast. cluster2 is young women with light LBP, so the intensity drop is not large but it belongs to the lowest generally.