

# **Cluster Analysis on Diagnosis-Related Group**

# Content

**Data Cleaning and Preparation**

**Cluster Exploration and Validation**

**Cluster Interpretation**

**Summary**

# 1

## Data Cleaning and Preparation

---



# 1

# Data Cleaning and Preparation

- Link DRG code and PCCR code with DRG and PCCR
- Filter DRGs between 20 and 977
- Merge filtered DRG to the Revenue Code file on UNIQ
- Exclude the low dollar value services (less than \$100)
- Sum all charges group by DRG, PCCR categories
- Cross tabulate with selected DRGs (in the row) and the mean value of the PCCR
- Combine the PCCR 3700 Operating Room & PCCR 4000 Anesthesiology
- Turn NA to 0

X	X3030.Angiocardiology	X3040.Audiology	X3050.Bacteriology.and.Microbiology	PCCR_OR_and_Anesth_Costs
1 Abortion w D&C, aspiration curettage...	0.000	0.0000	0.0000	5818.208
2 Abortion w/o D&C	0.000	0.0000	669.5000	330.990
3 Acute & subacute endocarditis w CC	0.000	0.0000	379.0000	1897.892
4 Acute & subacute endocarditis w MCC	0.000	0.0000	1728.0625	3437.576
5 Acute adjustment reaction & psychos...	0.000	0.0000	492.0000	171.870
6 Acute ischemic stroke w use of thro...	0.000	0.0000	0.0000	1747.490
7 Acute ischemic stroke w use of thro...	0.000	0.0000	0.0000	0.000
8 Acute ischemic stroke w use of thro...	0.000	0.0000	0.0000	0.000
9 Acute leukemia w/o major O.R. proce...	0.000	0.0000	0.0000	9124.578
10 Acute leukemia w/o major O.R. proce...	0.000	0.0000	0.0000	6690.758

687 rows , 55 columns



# 2

## Cluster Exploration and Validation

---



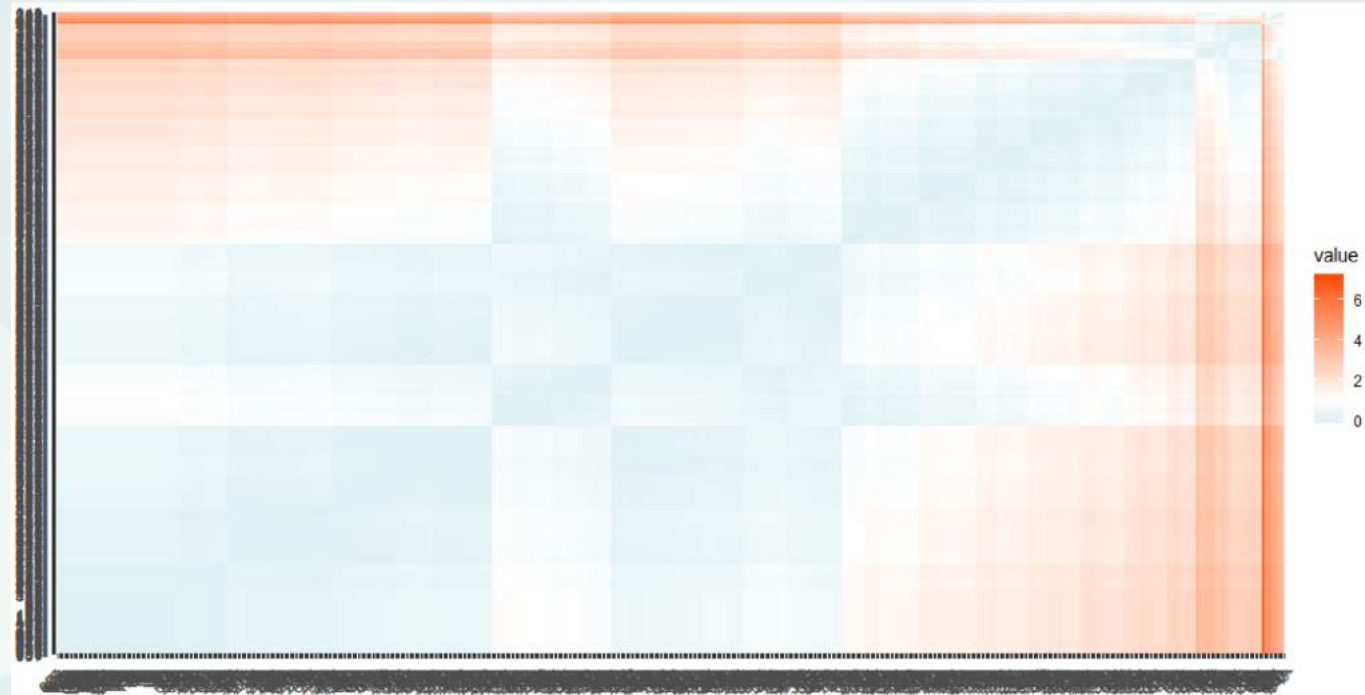
# 1 Data Exploration

## Step 1. Data Normalization

Find mean and standard deviation of the data.

## Step 2. Euclidean Distance Calculation and visualization

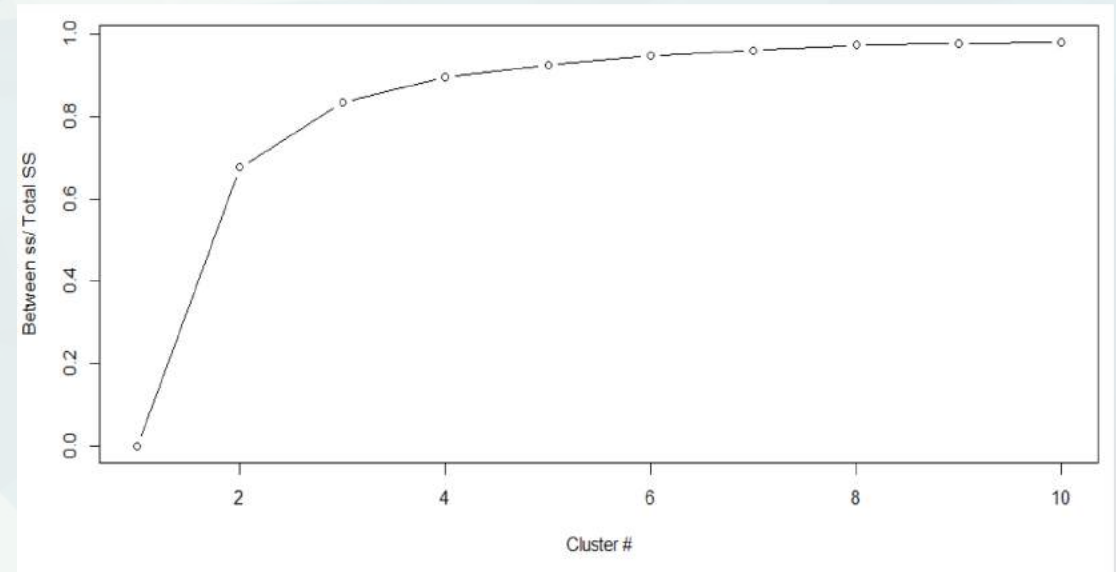
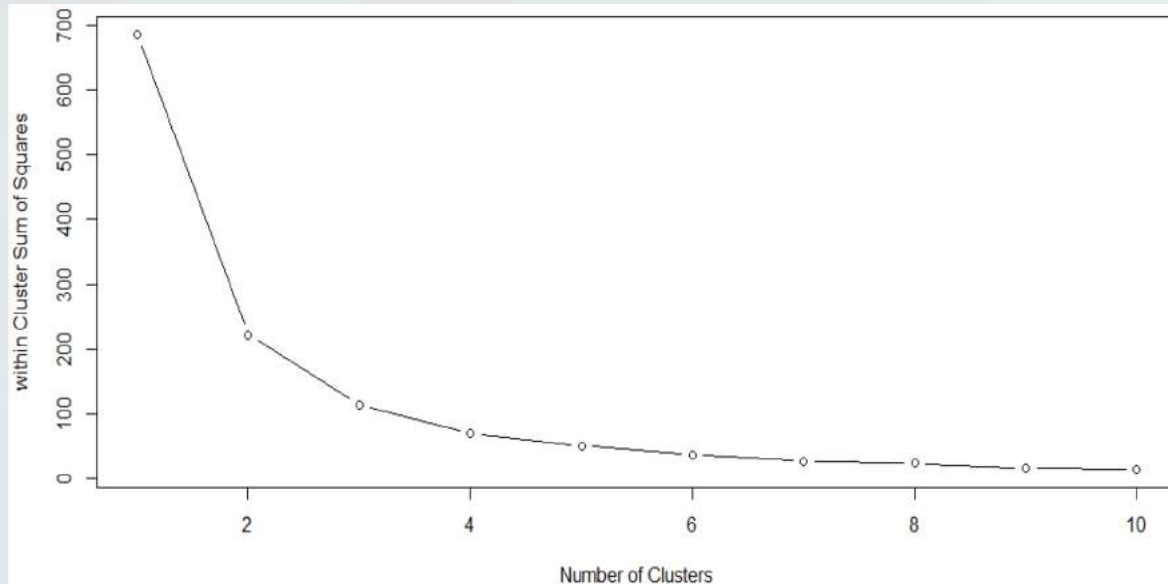
Most of the values are close to each other,  
Some values with larger distance.



## 2

# Only K-Means for Clustering

within Cluster Sum of Squares and between Cluster Sum of Squares under k from 1 to 10



3 or 4 is a better number for k

f-statistics value increases with cluster number increases

cluster	f_value
1	2 1429.75
2	3 1725.84
3	4 1978.02
4	5 2182.66

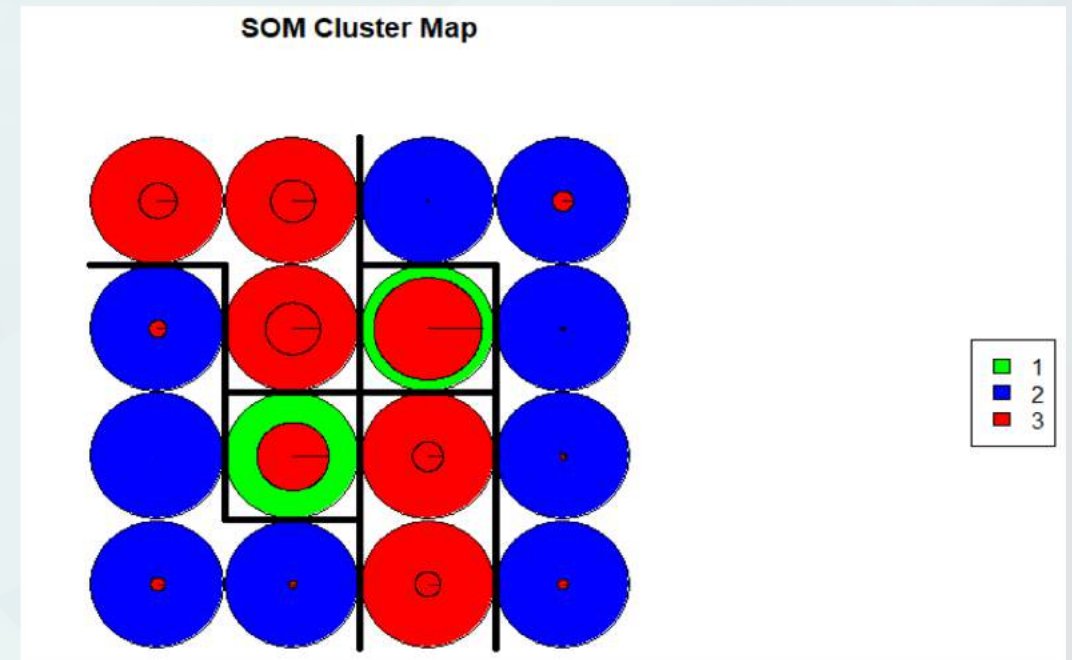
# 3

## SOM and K-Means for Clustering

SOM(Self Organizing Map):

Visualizing patterns by producing a 2 dimensional representation map.

1. Create 4 by 4 map and train data repeatedly fed into the model
1. Use k-means with  $k = 3$



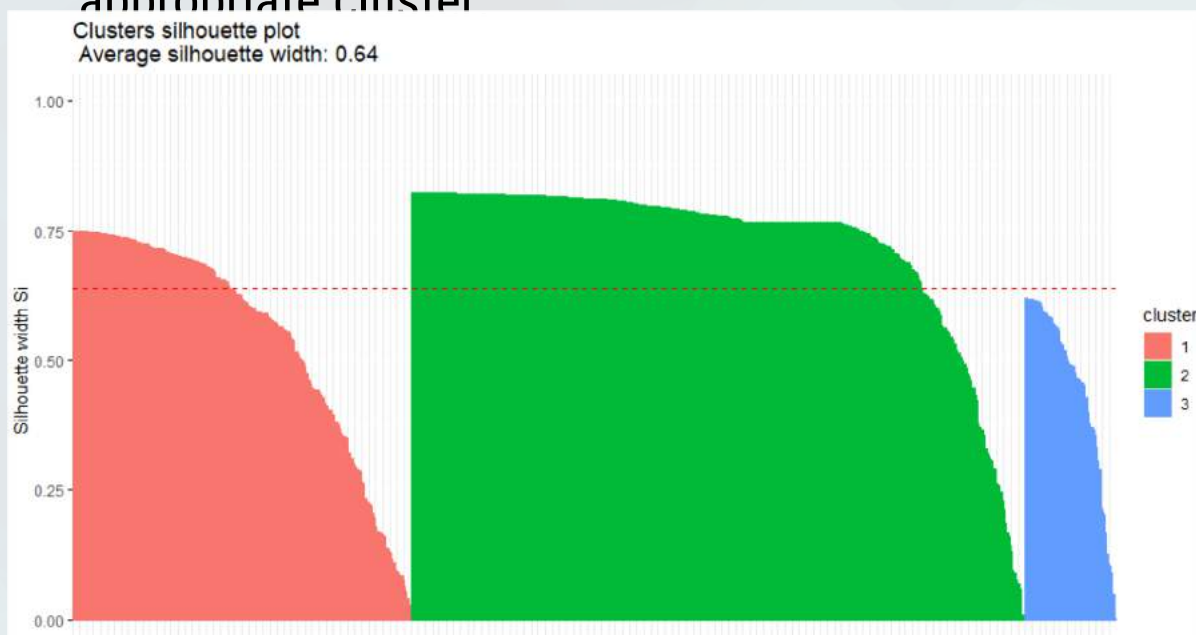


# 4

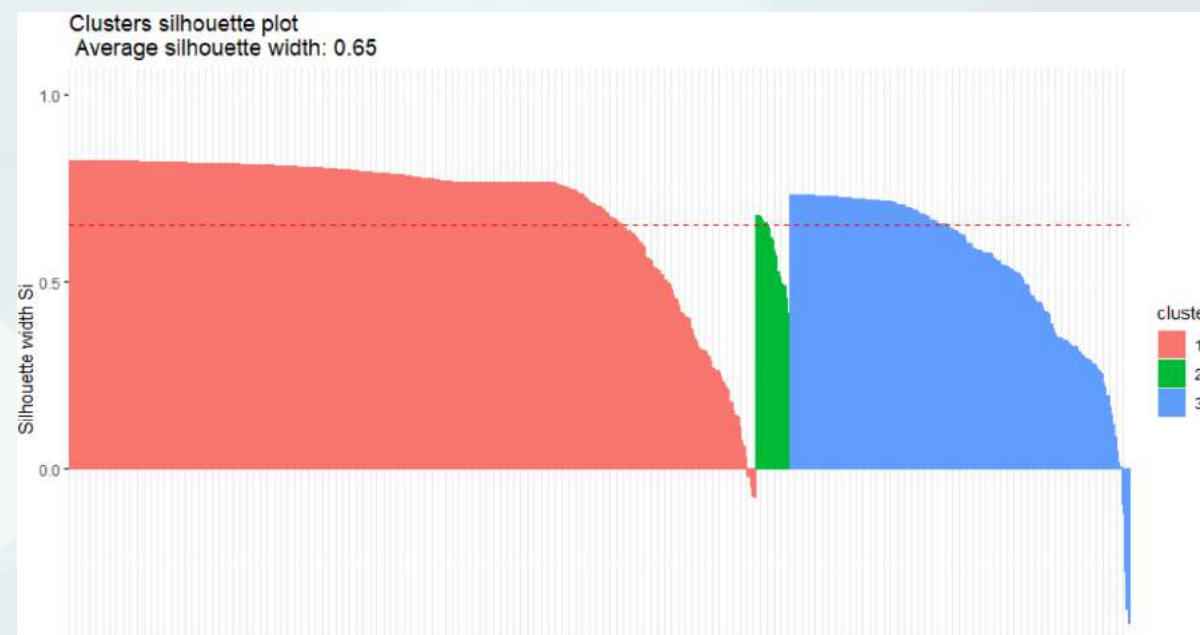
# Clustering Validation

## Silhouette Index

A measure of how similar an object is to its own cluster (cohesion) compared to other clusters (separation). If silhouette value is close to 1, sample is well-clustered and already assigned to a very appropriate cluster



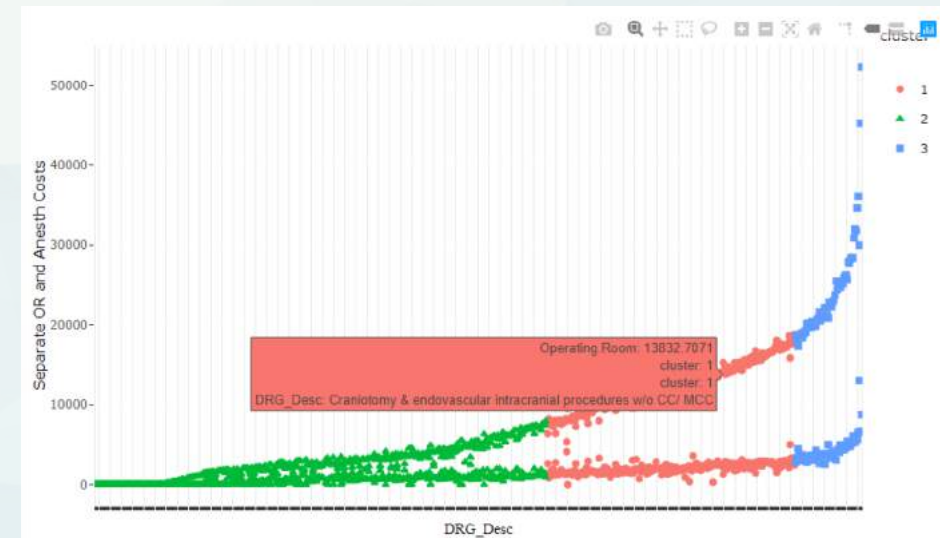
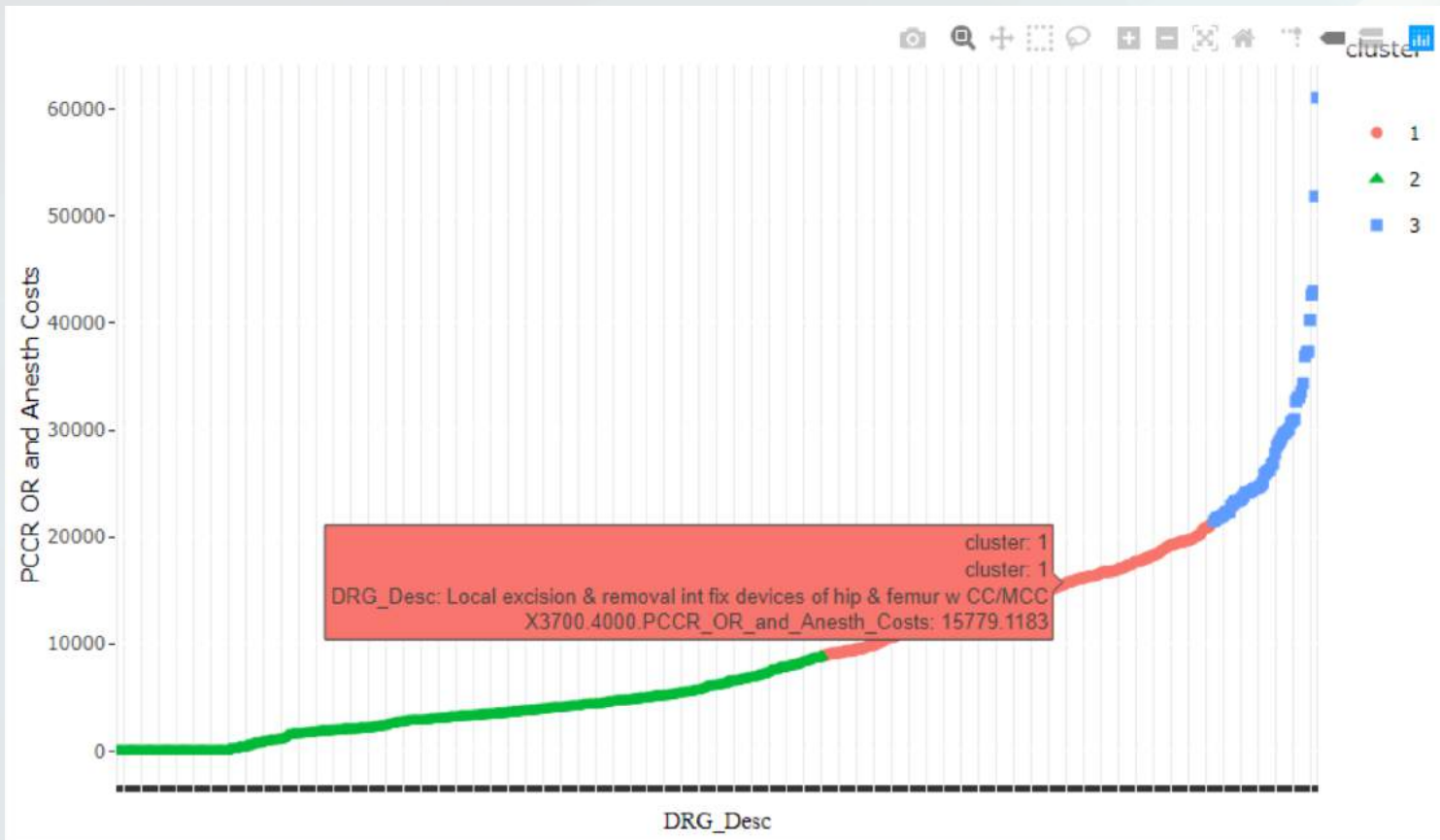
Method 1 Only K-Means



Method 2 SOM and K-Means

## 5

## Finalize Cluster



cluster	DRG_size	max_cost	min_cost
2	404	8,834.93	-
1	223	21,286.01	8,909.30
3	60	61,064.86	21,471.08

Choose the method 1 for our clustering

The cluster works well for the operating room and Anesthesiology separately too.

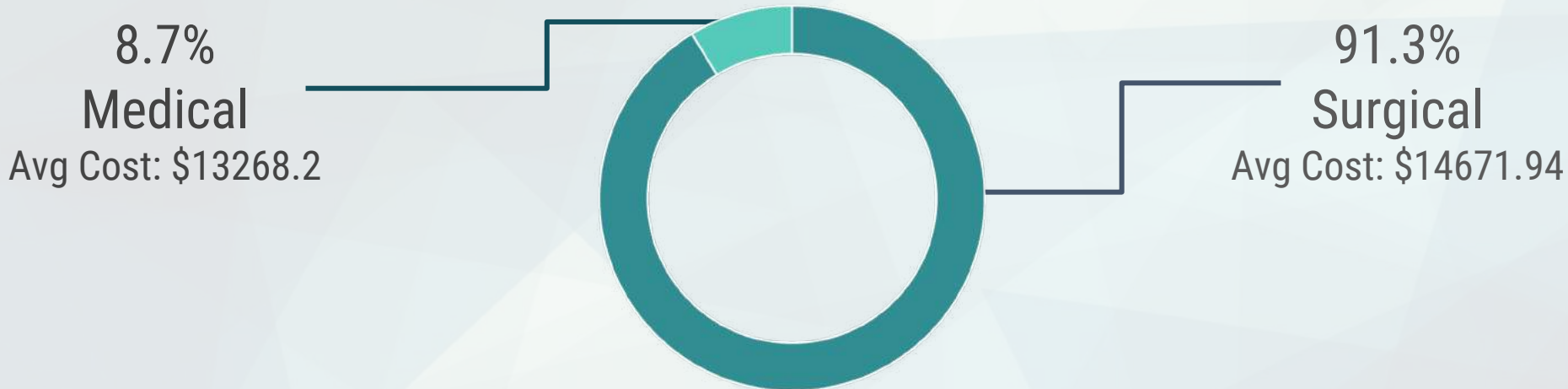
# 3

## Cluster 1 Mid Cost DRG

---

1

# Overview of Mid Cost DRG



## Medical DRG

No OR procedure is performed  
None or limited requirement for  
Operating room / Anesthesia

## Percentage of DRG types

Overall Avg Cost of this cluster:  
\$14008.7

## Surgical DRG

An OR procedure is performed  
May require the use of Operating  
Room / Anesthesia





# The Most Frequent Words in the Cluster







3

## Typical Organ System Covered by the Cluster

### 1. Muscular System

e.g. femur, joint, amputation, cranial

### 1. Urinary & Reproductive System

e.g. kidney, uterine, ureter

### 1. Integumentary system

e.g. skin, tissue

### 1. Serious Illness

e.g. malignancy, leukemia, cholecystectomy, appendectomy, lymphoma

# 4

## Time Length and Anesthesia Level of Some Surgery

### Muscular System

- Open Reduction Internal Fixation (OREF)
- Amputation Surgery

- The Surgery averagely takes 4-6 hours
- General or local anesthesia is required

### Urinary & Reproductive

- Endometriosis Surgery
- Urinary Dysfunction Surgery

- The Surgery averagely takes 1-2 hours
- Mostly using general anesthesia

### Integumentary System

- Mohs Surgery for skin Cancer

- The Surgery takes no longer than 4 hours
- Local anesthesia is mostly required

### Serious Illness

- Laparoscopic Cholecystectomy
- Appendix Removal Surgery

- The Surgery operation takes 1-3 hours
- General or local anesthesia is required

### Common Features

Source: Wikipedia, Mayo Clinic, etc.

The time length of surgical operation is not so long, that is, the operation is not too complicated

Both local anesthesia and general anesthesia are common

# 4

## Cluster 2 Low Cost DRG

---



2

## Cluster 2 mainly concentrated in medical treatment

12 out 24 MDCs contributed 80% of the cluster's total cost

4 MDCs costs above \$100K, and are top Primary Care visits, if serious, then proposed to surgical hospitals

Disease	Total Costs
MUSCULOSKELETAL	169,500
HEART & CIRCULATORY	147,813
DIGESTIVE	124,321
RESPIRATORY	119,773
BRAIN AND CNS	93,751
LIVER & PANCREAS	93,071
KIDNEY & URINARY	87,462
LYMPHATIC	58,212
PREGNANCY, CHILDBIRTH	55,089
ENDOCRINE	49,933
INFECTION	47,013
EAR, NOSE & THROAT	43,956
TTL	1,089,895

Disease	Total Costs	Min cost	Max cost	Average cost per DRG
MUSCULOSKELETAL	169,500	312	8,667	5,136
HEART & CIRCULATORY	147,813	161	7,594	3,438
DIGESTIVE	124,321	1,106	8,326	4,144
RESPIRATORY	119,773	161	7,530	3,743



# 5

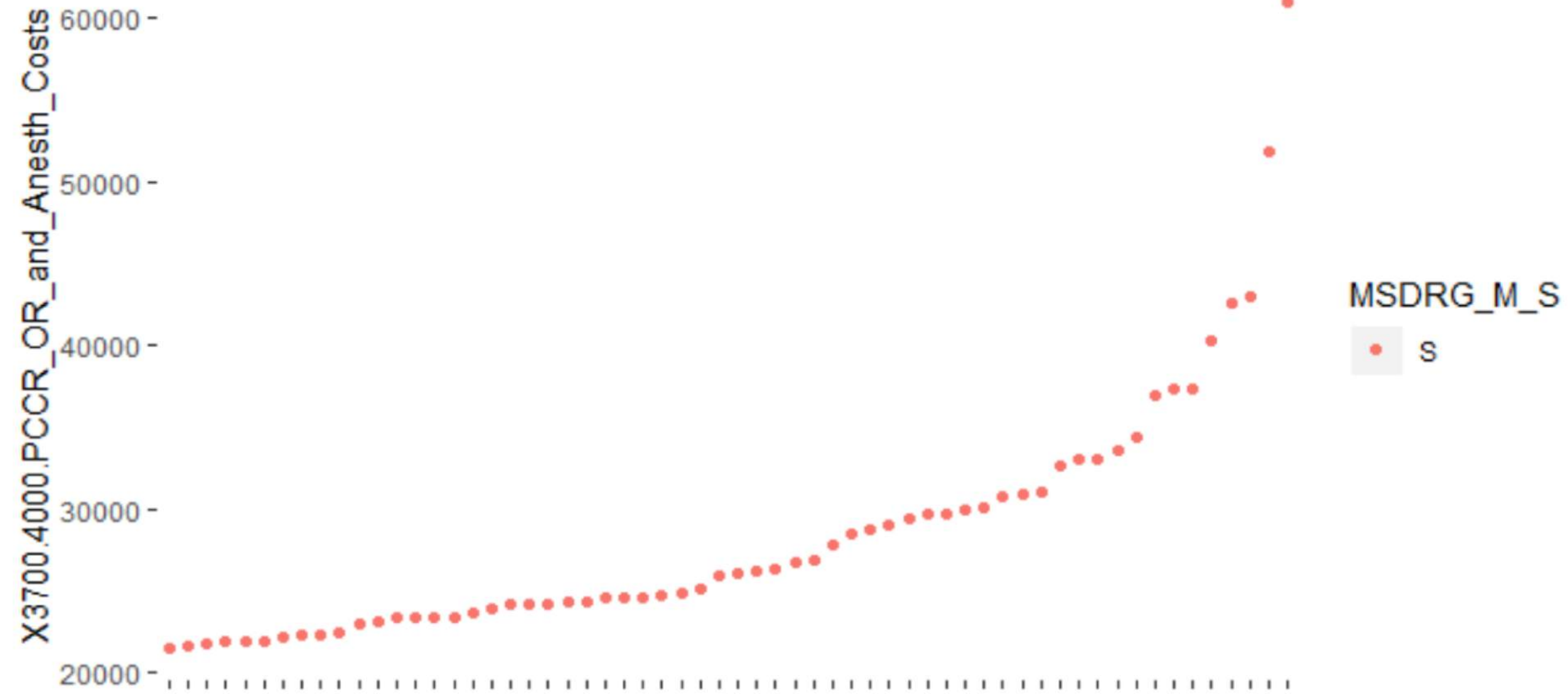
## Cluster 3 High Cost DRG

---



1

# Overview of High Cost DRG



2

## MDC in Cluster 3

MDC Percentage



- BRAIN AND CNS
- EAR, NOSE & THROAT
- HEART & CIRCULATORY
- DIGESTIVE
- LIVER & PANCREAS
- MUSCULOSKELETAL
- SKIN AND BREAST
- ENDOCRINE
- KIDNEY & URINARY
- FEMALE REPRODUCTIVE

# True Facts

- Heart disease is the leading cause of death for both men and women. More than half of the deaths due to heart disease in 2015 were in men.
- About 630,000 Americans die from heart disease each year—that's 1 in every 4 deaths.
- In the United States, someone has a heart attack every 40 seconds. Each minute, more than one person in the United States dies from a heart disease-related event.
- Heart disease costs the United States about \$200 billion each year.<sup>1</sup> This total includes the cost of healthcare services, medications, and lost productivity.

# 6

## SUMMARY

---



# 1

# Cluster Exploration



data exploration



2 ways to cluster the data



evaluate the goodness of clusters



choose and verify



02

01

03

Lorem Ipsum is simply dummy text of the printing and typesetting industry. Lorem Ipsum has been the industry's standard dummy text ever since the 1500s, when an unknown printer took a galley of type and scrambled it to make a type specimen book. It has survived not only five centuries, but also the leap into electronic typesetting, remaining essentially unchanged. It was popularised in the 1960s with the release of Letraset sheets containing Lorem Ipsum passages, and more recently with desktop publishing software.

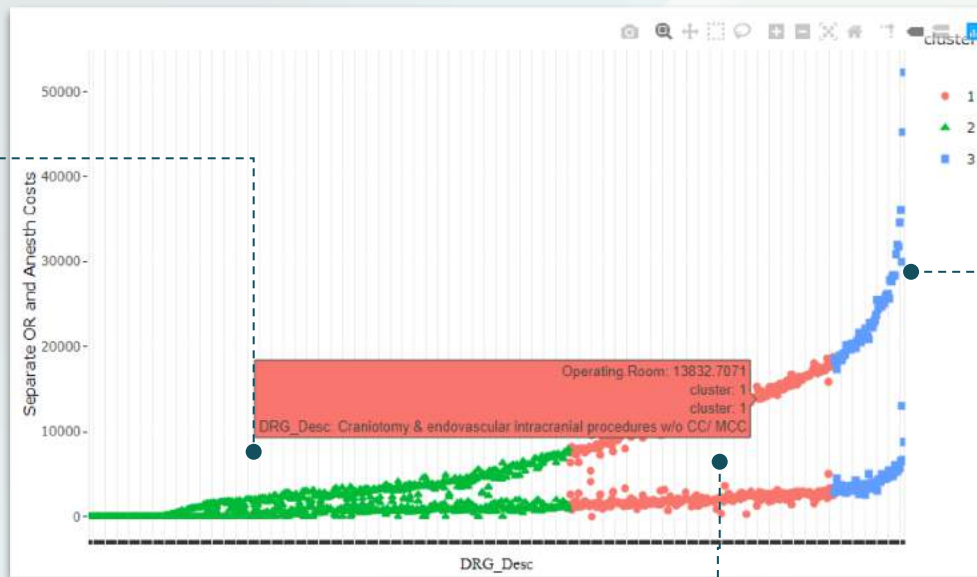


# 2

## Cluster Interpretation

### Cluster 2: Low Cost DRG

- Mainly concentrated on medical treatment
- The top 4 MDCs are the top Primary Care visits



### Cluster 1: Mid Cost DRG

- Main cost is surgical treatment
- Moderately complex surgical treatments on non-heart and brain systems

### Cluster 3: High Cost DRG

- all elements in this cluster are under the surgical category
- Major parts: heart & Circulatory ; musculoskeletal treatments