# ML 1000 Assignment 2

by Anupama r.k, Queenie Tsang, Crystal (Yunan) Zhu

# 21/02/2021

# To do list:

- Add Pie charts! by sub\_category, region # (done)
- Create a Month variable to see the change of sales/profits by month?
- bar charts of profits/sales by region #(done)
- Output the characteristics of the orders with the highest and lowest profits/sales e.g. what made the order? when? bought what product? in which city/state/region? Any discount?
- relationship between discount & sales, discount & profits, sales & profits, and the role of region?
- from someone's analysis there is no significant change between the four discount categories when it comes to Sales
- sales/profits by month, rather than by date? color by region?

### Abstract

Anomaly detection or Outlier detection identifies data points, events or observations that deviate from dataset's normal behavior. Anomalous data indicate critical incidents or potential opportunities. In order to take advantage of opportunities or fix costly problems anomaly detection has to be done in real time. Unsupervised machine learning models can be used to automate anomaly detection. Unsupervised anomaly detection algorithms scores data based on intrinsic properties of the dataset. Distances and densities are used to give an estimation what is normal and what is an outlier. Anomaly detection monitor is a tool developed for an online retailer to check product quality issues like profit opportunities and sales glitches. The application is built using R and Shinyapp following CRISP-DM framework.

# **Business Case**

# **Objective**

Detect point anomalies from superstore dataset using K-NN and clustering methods.

### Data Understanding

US Superstore dataset is sourced from US uperstore dataset . The dataset have online orders for Superstores in U.S. from 2014-2018. Tableau community is the owner of the dataset. The dataset has 9994 records and 21 attributes.

### Import data

```
superstore<- read_excel("US_Superstore_data.xls")</pre>
## Warning in read_fun(path = enc2native(normalizePath(path)), sheet_i = sheet, :
## Coercing text to numeric in L2236 / R2236C12: '05408'
## Warning in read_fun(path = enc2native(normalizePath(path)), sheet_i = sheet, :
## Coercing text to numeric in L5276 / R5276C12: '05408'
## Warning in read_fun(path = enc2native(normalizePath(path)), sheet_i = sheet, :
## Coercing text to numeric in L8800 / R8800C12: '05408'
## Warning in read_fun(path = enc2native(normalizePath(path)), sheet_i = sheet, :
## Coercing text to numeric in L9148 / R9148C12: '05408'
## Warning in read_fun(path = enc2native(normalizePath(path)), sheet_i = sheet, :
## Coercing text to numeric in L9149 / R9149C12: '05408'
## Warning in read_fun(path = enc2native(normalizePath(path)), sheet_i = sheet, :
## Coercing text to numeric in L9150 / R9150C12: '05408'
## Warning in read_fun(path = enc2native(normalizePath(path)), sheet_i = sheet, :
## Coercing text to numeric in L9388 / R9388C12: '05408'
## Warning in read_fun(path = enc2native(normalizePath(path)), sheet_i = sheet, :
## Coercing text to numeric in L9389 / R9389C12: '05408'
## Warning in read_fun(path = enc2native(normalizePath(path)), sheet_i = sheet, :
## Coercing text to numeric in L9390 / R9390C12: '05408'
## Warning in read_fun(path = enc2native(normalizePath(path)), sheet_i = sheet, :
## Coercing text to numeric in L9391 / R9391C12: '05408'
## Warning in read_fun(path = enc2native(normalizePath(path)), sheet_i = sheet, :
## Coercing text to numeric in L9743 / R9743C12: '05408'
data_superstore
```

Table 1: Dataset description

Attribute	Data Type	Description
Row ID	numeric	row number
Order ID	character	unique order number
Order Date	numeric	order placed date
Ship Date	numeric	order shipping date
Ship Mode	character	shipping mode of order
Customer ID	character	unique customer id for order

Attribute	Data Type	Description		
Customer Name character		name of customer		
Segment	character	section of product		
Country	character	country based on order		
City	character	city based on order		
State	character	state based on order		
Postal Code	numeric	pin code		
Region	character	region based on order		
Product ID	character	product id of product		
Category	character	category of product		
Sub-Category	character	sub-category of product		
Product Name	character	name of product		
Sales	numeric	selling price of product		
Quantity	numeric	order quantity		
Discount	numeric	discount on product		
Profit	numeric	profit from product		

```
## [1] "i..Row.ID-0 missing values"
                                         "Order.ID-0 missing values"
   [3] "Order.Date-0 missing values"
                                         "Ship.Date-0 missing values"
##
## [5] "Ship.Mode-0 missing values"
                                         "Customer.ID-0 missing values"
  [7] "Customer.Name-O missing values"
                                         "Segment-0 missing values"
  [9] "Country-O missing values"
                                         "City-O missing values"
##
## [11] "State-0 missing values"
                                         "Postal.Code-0 missing values"
## [13] "Region-0 missing values"
                                         "Product.ID-0 missing values"
## [15] "Category-O missing values"
                                         "Sub.Category-O missing values"
## [17] "Product.Name-0 missing values"
                                         "Sales-0 missing values"
## [19] "Quantity-0 missing values"
                                         "Discount-O missing values"
## [21] "Profit-0 missing values"
                                         "diff_in_days-0 missing values"
```

Get a general idea of the data set.

```
length(unique(data$Customer.ID))
```

## [1] 793

```
#793 unique customer IDs
length(unique(data$Customer.Name))
```

## [1] 793

```
#793 unique customer names - drop one of these two vars
length(unique(data$Order.Date))
```

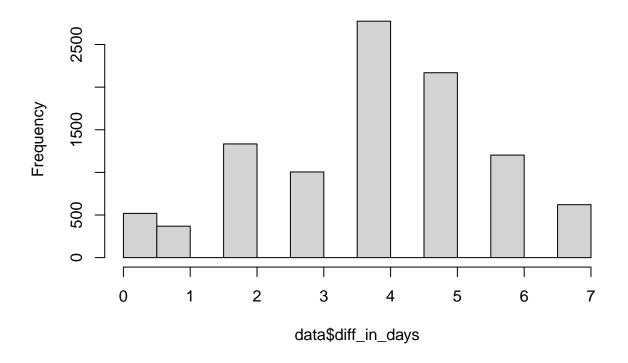
## [1] 1237

```
#1237 unique order dates
length(unique(data$Ship.Date))
```

```
## [1] 1334
#1334 unique ship dates - more unique ship dates than order dates - orders made on the same day were sh
length(unique(data$Segment))
## [1] 3
unique(data$Segment)
## [1] "Consumer"
                     "Corporate" "Home Office"
#"Consumer" "Corporate" "Home Office"
unique(data$Country)
## [1] "United States"
#all are from US - could drop this variable due to no-variation introduced by it
length(unique(data$City))
## [1] 531
#531 different cities
length(unique(data$State))
## [1] 49
#49 states
length(unique(data$Postal.Code))
## [1] 631
#631 postal code - 793 unique customer IDs - some customers live very close!
unique(data$Region)
## [1] "South"
                 "West"
                           "Central" "East"
#only 4 regions
unique(data$Category)
## [1] "Furniture"
                         "Office Supplies" "Technology"
```

```
#only 3 categories - "Furniture" "Office Supplies" "Technology"
length(unique(data$Sub.Category))
## [1] 17
unique(data$Sub.Category)
## [1] "Bookcases"
                      "Chairs"
                                    "Labels"
                                                  "Tables"
                                                                "Storage"
## [6] "Furnishings" "Art"
                                    "Phones"
                                                                "Appliances"
                                                  "Binders"
                                                                "Supplies"
## [11] "Paper"
                      "Accessories" "Envelopes"
                                                  "Fasteners"
## [16] "Machines"
                     "Copiers"
#17 sub-categories
length(unique(data$Product.Name))
## [1] 1850
#1850 product names
length(unique(data$Product.ID))
## [1] 1862
#1862 product IDs - potential redundant variables!
hist(data$diff_in_days)
```

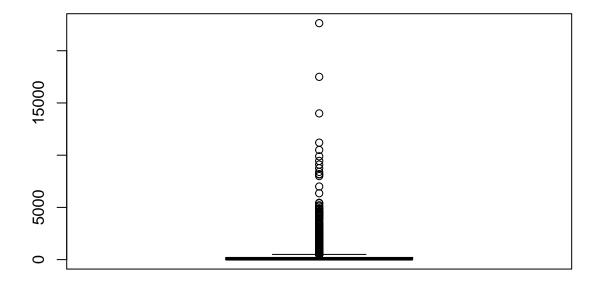
# Histogram of data\$diff\_in\_days



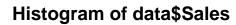
#The time difference between order date and ship date typically takes 4 days.
summary(data\$Sales)

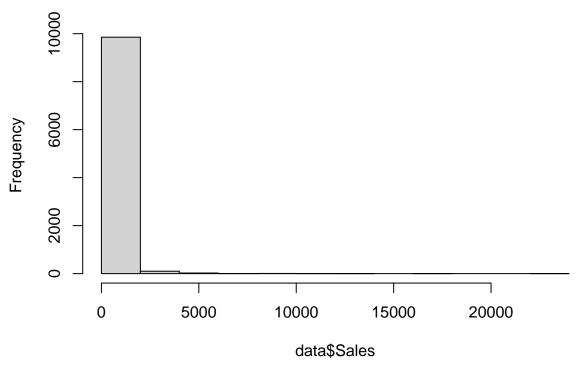
```
## Min. 1st Qu. Median Mean 3rd Qu. Max.
## 0.444 17.280 54.490 229.858 209.940 22638.480
```

boxplot(data\$Sales)



hist(data\$Sales)

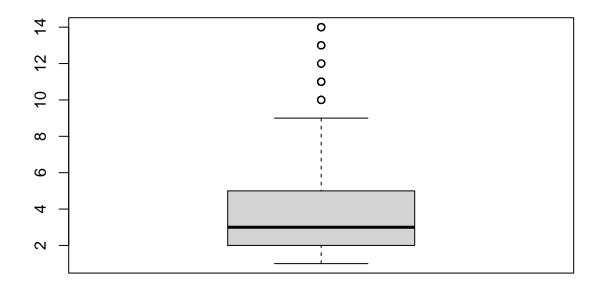




```
#a large amount of orders with very small Sales!
summary(data$Quantity)

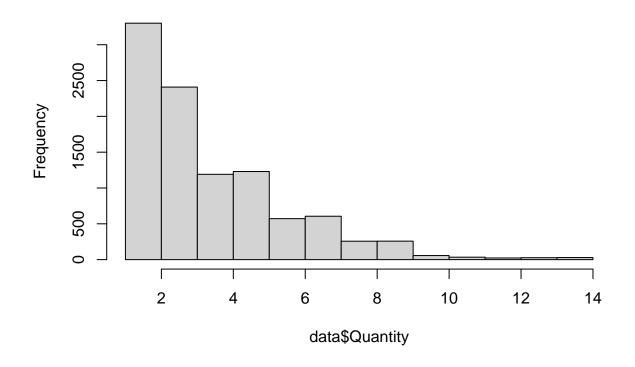
## Min. 1st Qu. Median Mean 3rd Qu. Max.
## 1.00 2.00 3.00 3.79 5.00 14.00

boxplot(data$Quantity)
```



#not many outliers - the #of products in each order is stable?
hist(data\$Quantity)

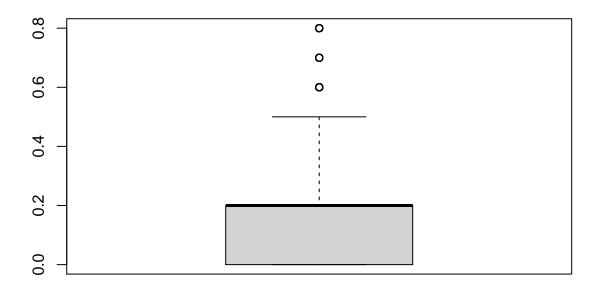
# Histogram of data\$Quantity



```
#very skewed distribution - most of the orders have small #of items
summary(data$Discount)
```

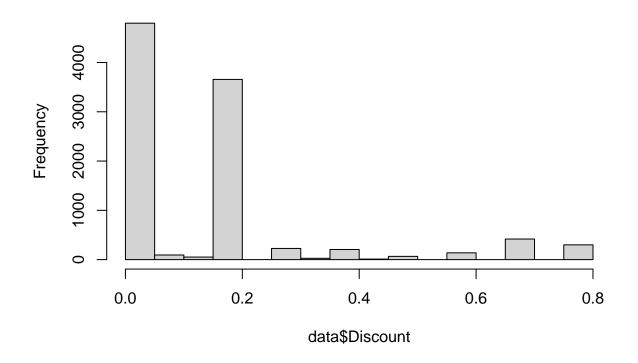
```
## Min. 1st Qu. Median Mean 3rd Qu. Max.
## 0.0000 0.0000 0.2000 0.1562 0.2000 0.8000
```

boxplot(data\$Discount)



#a strange looking box dataplot? - median & 3rd quantile are the same (0.2) - not many orders have high hist(data\$Discount)

# Histogram of data\$Discount



#most of the orders were placed without any discounts or with 20% off
summary(data\$Profit)

## Min. 1st Qu. Median Mean 3rd Qu. Max.

29.364

8399.976

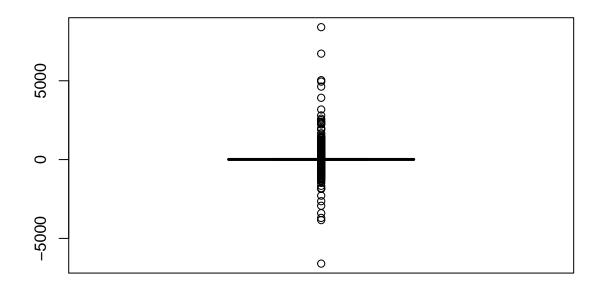
28.657

8.666

boxplot(data\$Profit)

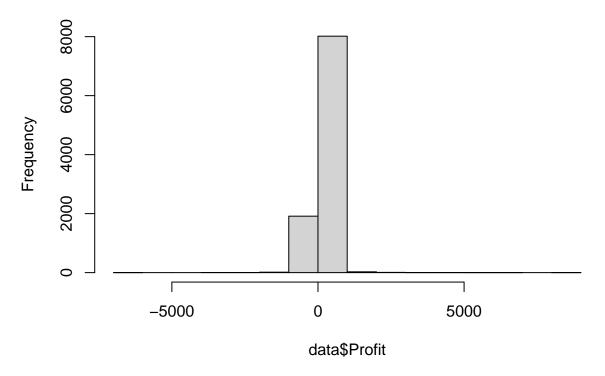
1.729

## -6599.978



#most of the profits are outside of the box - but most of them clustered close to the box(not with so e hist(dataProfit)

# Histogram of data\$Profit



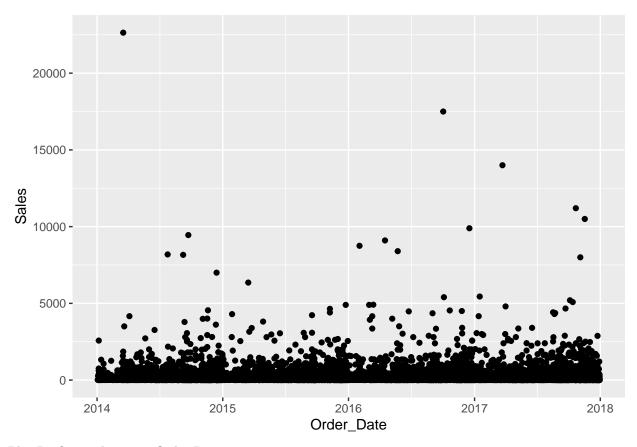
#most of the orders have profits ~1000 (or ~800?), and ~ -800

Remove the dot in the column names and replace with "\_\_" to make variable names easier to handle:

```
##
    [1] "i__Row_ID"
                         "Order_ID"
                                          "Order_Date"
                                                            "Ship_Date"
       "Ship_Mode"
                                          "Customer_Name"
                                                           "Segment"
                         "Customer_ID"
    [9]
        "Country"
                         "City"
                                          "State"
                                                            "Postal_Code"
        "Region"
                         "Product_ID"
                                                            "Sub_Category"
   [13]
                                          "Category"
        "Product_Name"
                         "Sales"
                                          "Quantity"
                                                            "Discount"
   [17]
## [21] "Profit"
                         "diff_in_days"
```

# **Exploratory Data Analysis**

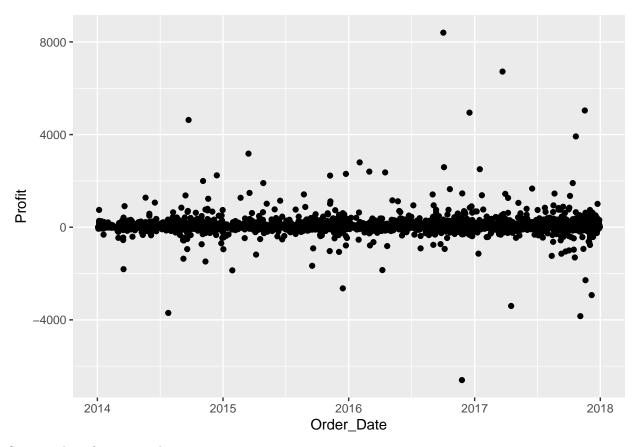
Plot Sales in relation to Order Date:



Plot Profit in relation to Order Date:

```
ggplot(data = data) +
geom_point(mapping = aes(x = Order_Date, y = Profit), xlab="Order Date", ylab="Profit")
```

## Warning: Ignoring unknown parameters: xlab, ylab



Some outliers for certain days

```
table(data$'Sub_Category')
```

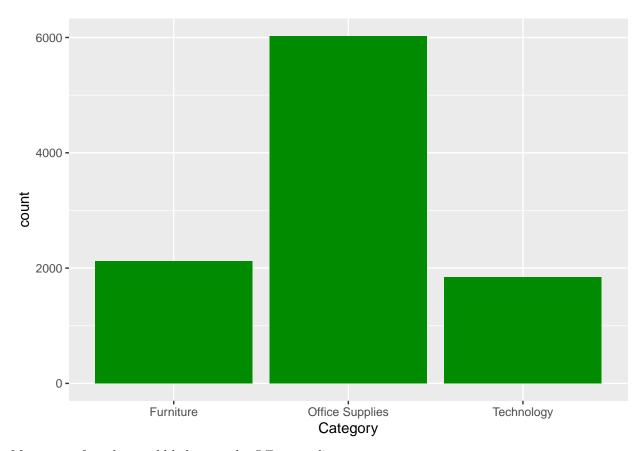
##						
##	Accessories	Appliances	Art	Binders	Bookcases	Chairs
##	775	466	796	1523	228	617
##	Copiers	Envelopes	Fasteners	Furnishings	Labels	Machines
##	68	254	217	957	364	115
##	Paper	Phones	Storage	Supplies	Tables	
##	1370	889	846	190	319	

look at the time range for these transactions, ie. start date for Order\_Date column:

```
## Min. 1st Qu. Median Mean 3rd Qu. Max.
## "2014-01-03" "2015-05-23" "2016-06-26" "2016-04-30" "2017-05-14" "2017-12-30"
#[1] min "2014-01-03", max "2017-12-30"
```

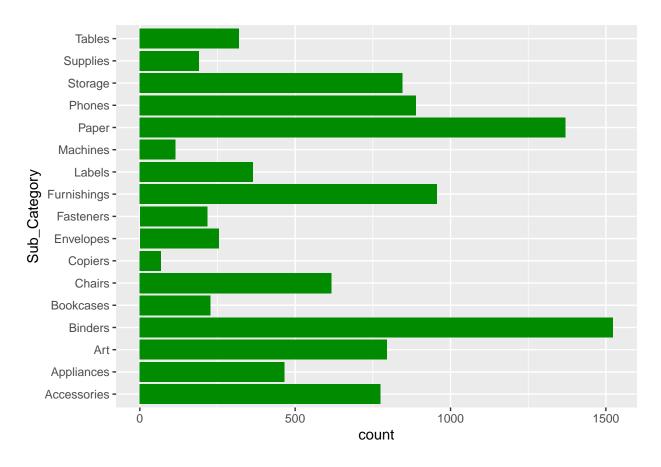
Basically this dataset covers transactions ranging from 2014-01-03 to 2017-12-30.

```
ggplot(data = data) +
geom_bar(mapping = aes(x = Category),fill="green4")
```



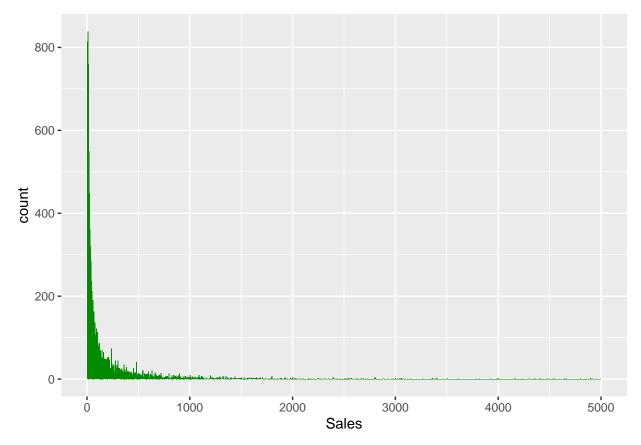
Most type of products sold belong to the Office supplies category.

```
ggplot(data = data) +
geom_bar(mapping = aes(y = 'Sub_Category'), fill="green4")
```



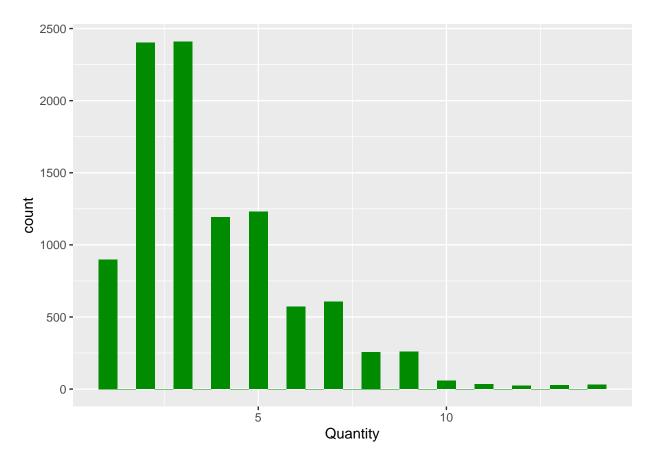
```
ggplot(data = data, mapping = aes(x = Sales)) +
    xlim(0, 5000) +
    geom_histogram(binwidth = 5,fill="green4")
```

- ## Warning: Removed 19 rows containing non-finite values (stat\_bin).
- ## Warning: Removed 2 rows containing missing values (geom\_bar).

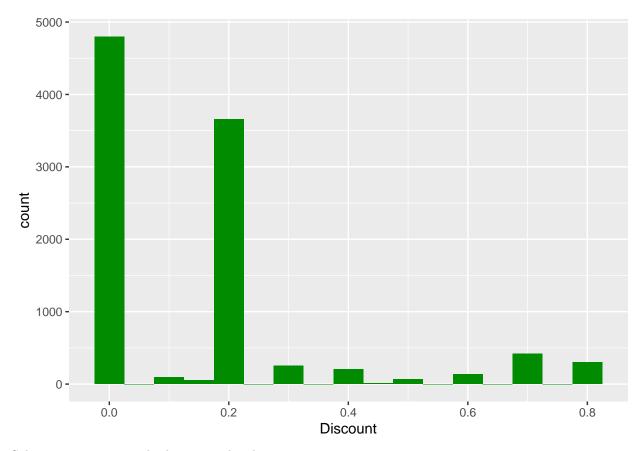


Most sales are very few items (<500).

```
ggplot(data = data, mapping = aes(x = Quantity)) +
geom_histogram(binwidth = 0.5,fill="green4")
```



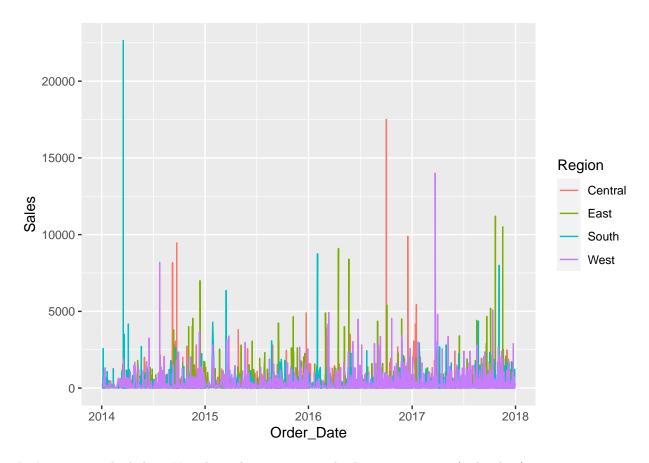
## Warning: Ignoring unknown parameters: xlab



Sales transactions mostly do not involve discounts.

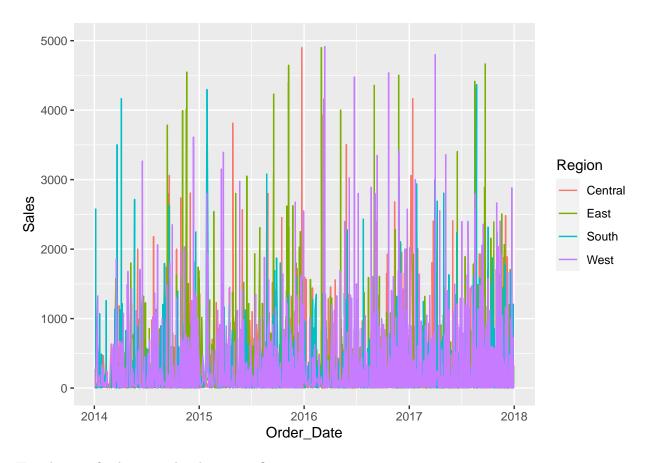
Visualise sales transactions by Region over time (order date).

```
ggplot(data, aes(Order_Date, Sales,color=Region)) +
    geom_line()
```



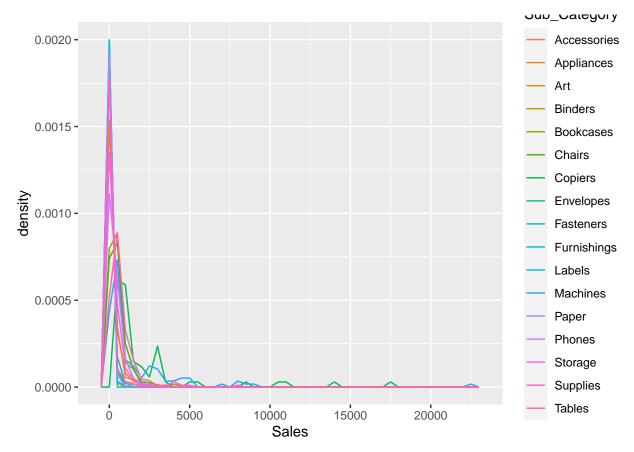
Let's zoom in a little bit - Visualise sales transactions by Region over time (order date).

```
ggplot(data, aes(Order_Date, Sales,color=Region)) +
   geom_line() +
   ylim(0,5000)
```



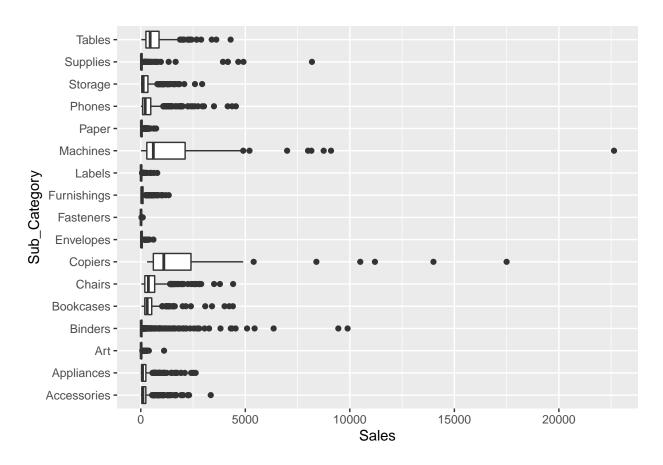
How does profit change with sub-category?

```
#density plot where the count is standardized, area under each frequency is 1
ggplot(data = data, mapping = aes(x = Sales, y = ..density..)) +
geom_freqpoly(mapping = aes(colour = Sub_Category), binwidth = 500)
```

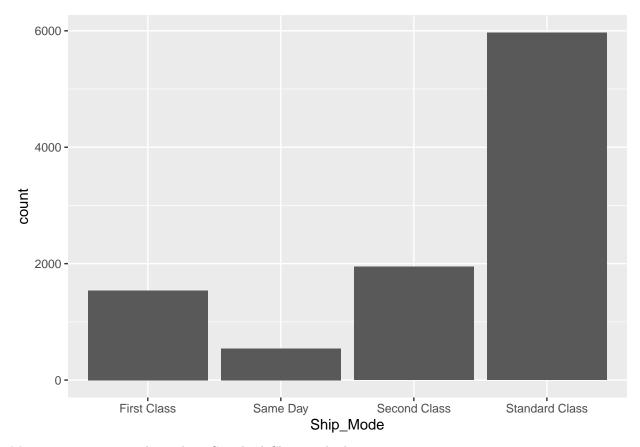


It looks like some categories of items ie. supplies or accessories have negative sales values. How does sales vary across sub category?

```
ggplot(data = data, mapping = aes(x = Sales, y = 'Sub_Category')) +
geom_boxplot()
```



```
ggplot(data =data, mapping = aes(x = Ship_Mode)) +
  geom_bar()
```

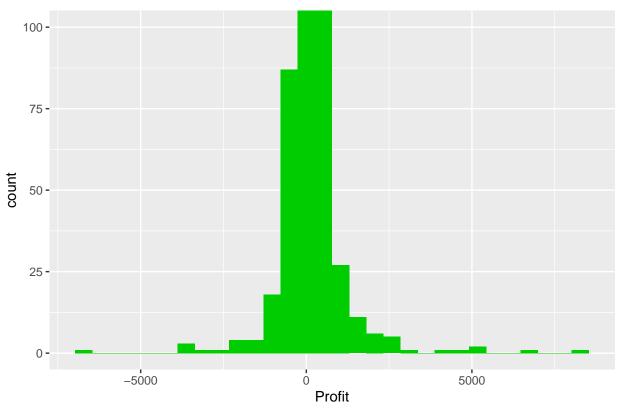


Most transactions are shipped via Standard Class method.

```
ggplot(data)+
geom_histogram(mapping=aes(x=Profit),fill="green3")+
coord_cartesian(ylim = c(0, 100))+
labs(title=" Profit Distribution")
```

## 'stat\_bin()' using 'bins = 30'. Pick better value with 'binwidth'.

# **Profit Distribution**



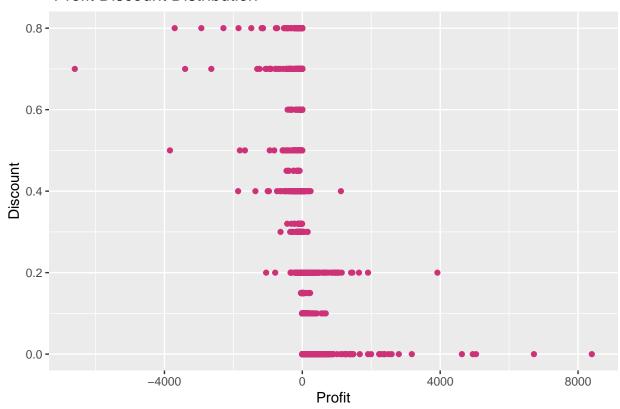
```
ggplot(data)+
geom_histogram(mapping=aes(x=Sales),fill="sienna3")+
coord_cartesian(ylim = c(0, 100))+labs(title=" Sales Distribution")
```

## 'stat\_bin()' using 'bins = 30'. Pick better value with 'binwidth'.

# Sales Distribution 10075250005000 10000 Sales

```
ggplot(data) +
geom_point(mapping = aes(x = Profit, y = Discount),colour="violetred3")+
labs(title=" Profit Discount Distribution")
```

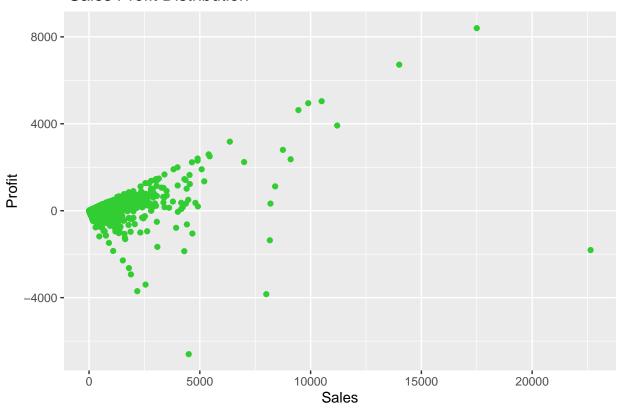
# **Profit Discount Distribution**



### Sales Profit

```
ggplot(data) +
geom_point(mapping = aes(x = Sales, y = Profit),colour="limegreen")+
labs(title=" Sales Profit Distribution")
```

# Sales Profit Distribution



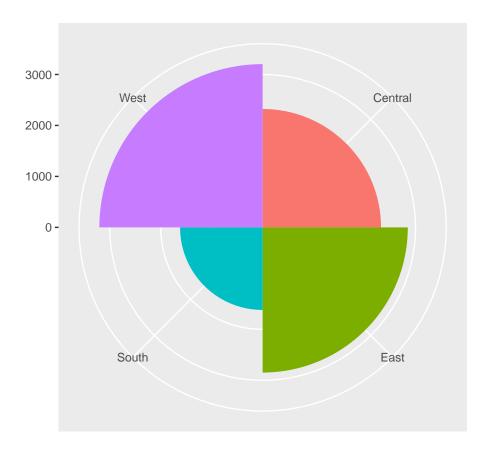
```
#product name and product id mismatch
data %>%
  distinct(Product_Name,Product_ID) %>%
  group_by(Product_ID) %>%
  filter(n()>1) %>%
  select(Product_ID)
```

```
## # A tibble: 64 x 1
## # Groups:
              Product_ID [32]
##
     Product_ID
##
      <chr>
  1 FUR-FU-10004848
##
## 2 FUR-CH-10001146
## 3 OFF-BI-10004654
## 4 FUR-CH-10001146
## 5 OFF-PA-10002377
## 6 OFF-AR-10001149
## 7 OFF-PA-10000659
## 8 TEC-MA-10001148
## 9 FUR-FU-10004017
## 10 TEC-AC-10003832
## # ... with 54 more rows
```

# #total category and subcategory

```
count_category<-unique(data$Category)</pre>
length(count_category)
## [1] 3
count_subcategory<-unique(data$Sub_Category)</pre>
length(count_subcategory)
## [1] 17
data %>%
 distinct(Category, Sub_Category)
##
             Category Sub_Category
## 1
            Furniture
                         Bookcases
## 2
            Furniture
                             Chairs
## 3 Office Supplies
                             Labels
## 4
            Furniture
                             Tables
## 5 Office Supplies
                            Storage
            Furniture Furnishings
## 6
## 7 Office Supplies
                                Art
## 8
           Technology
                             Phones
## 9 Office Supplies
                            Binders
## 10 Office Supplies
                        Appliances
## 11 Office Supplies
                              Paper
## 12
           Technology Accessories
## 13 Office Supplies
                         Envelopes
## 14 Office Supplies
                         Fasteners
## 15 Office Supplies
                           Supplies
## 16
           Technology
                           Machines
## 17
           Technology
                            Copiers
superstore_sales<-data %>%
                  select(Order_Date,Sales)
superstore_sales<-as_tibble(superstore_sales)</pre>
Transactions by region:
bar <- ggplot(data = data) +</pre>
  geom_bar(
    mapping = aes(x = Region, fill = Region),
    show.legend = FALSE,
    width = 1
 ) +
  theme(aspect.ratio = 1) +
 labs(x = NULL, y = NULL)
```

bar + coord\_polar()

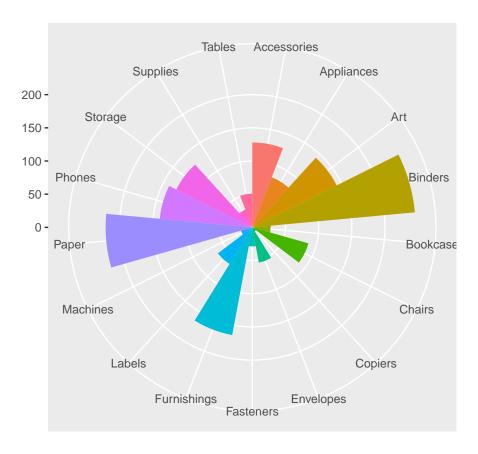


The above chart shows proportions of transactions from the different regions.

```
#Extracting the rows for South region, and sub-categories:
South <- data %>%
    select(Region, Sub_Category) %>%
    filter(Region == "South")

bar <- ggplot(data = South) +
    geom_bar(
        mapping = aes(x = Sub_Category, fill = Sub_Category),
        show.legend = FALSE,
        width = 1
    ) +
    theme(aspect.ratio = 1) +
    labs(x = NULL, y = NULL)

bar + coord_polar()</pre>
```

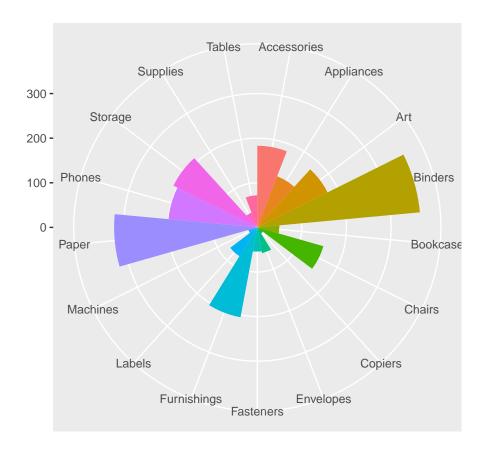


In the South, most transactions are Binders, Paper, or Furnishings.

```
#Extracting the rows for Central region, and sub-categories:
Central <- data %>%
    select(Region, Sub_Category) %>%
    filter(Region == "Central")

bar <- ggplot(data = Central) +
    geom_bar(
        mapping = aes(x = Sub_Category, fill = Sub_Category),
        show.legend = FALSE,
        width = 1
    ) +
    theme(aspect.ratio = 1) +
    labs(x = NULL, y = NULL)

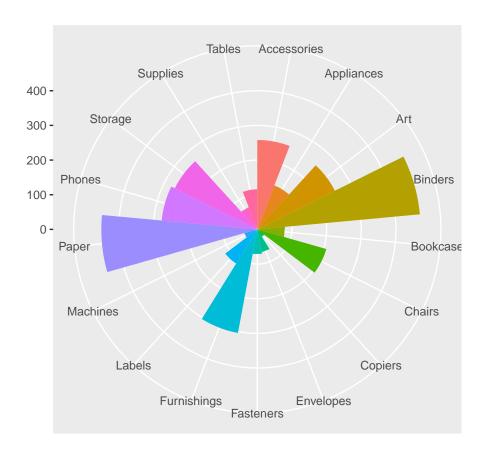
bar + coord_polar()</pre>
```



```
#Extracting the rows for West region, and sub-categories:
West <- data %>%
    select(Region, Sub_Category) %>%
    filter(Region == "West")

bar <- ggplot(data = West) +
    geom_bar(
        mapping = aes(x = Sub_Category, fill = Sub_Category),
        show.legend = FALSE,
        width = 1
    ) +
    theme(aspect.ratio = 1) +
    labs(x = NULL, y = NULL)

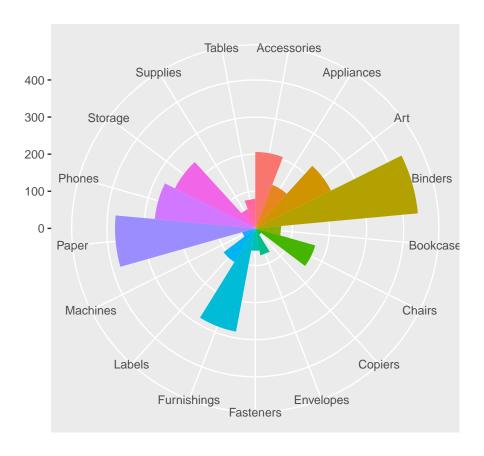
bar + coord_polar()</pre>
```



```
#Extracting the rows for East region, and sub-categories:
East <- data %>%
    select(Region, Sub_Category) %>%
    filter(Region == "East")

bar <- ggplot(data = East) +
    geom_bar(
        mapping = aes(x = Sub_Category, fill = Sub_Category),
        show.legend = FALSE,
        width = 1
    ) +
    theme(aspect.ratio = 1) +
    labs(x = NULL, y = NULL)

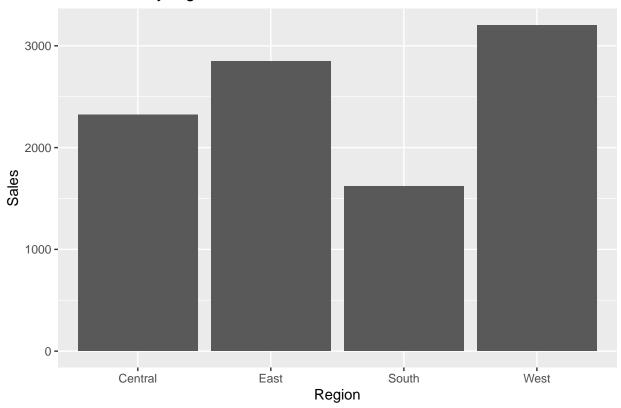
bar + coord_polar()</pre>
```



• bar charts of profits/sales by region

```
ggplot(data = data) +
  geom_bar(mapping = aes(x = Region, fill = Sales)) +
  ggtitle("Total Sales by region") +
  ylab("Sales")
```

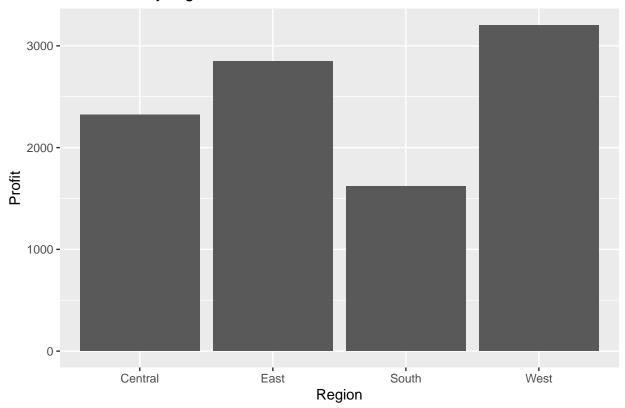
# Total Sales by region



Total sales per region.

```
ggplot(data = data) +
  geom_bar(mapping = aes(x = Region, fill = Profit))+
  ggtitle("Total Profit by region")+
  ylab("Profit")
```

# Total Profit by region



Look at relationship between numeric variables:

```
#subset the numeric variables:
numeric_vars<- c("Sales", "Quantity", "Discount", "Profit", "diff_in_days")
num_data <- data[numeric_vars]</pre>
```

We'll use a correlation matrix to look at the relationship between numeric variables:

```
cor(num_data)
```

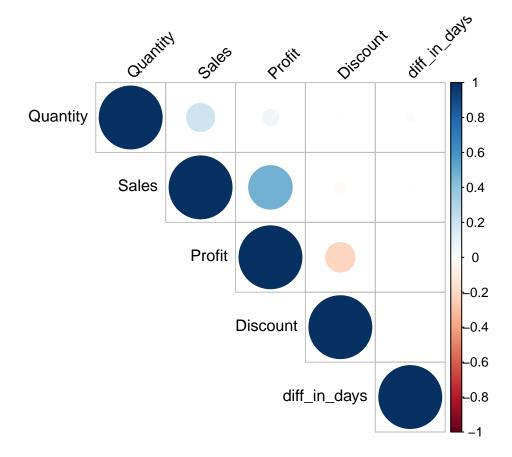
```
## Sales Quantity Discount Profit diff_in_days
## Sales 1.00000000 0.20079477 -0.0281901242 0.479064350 -0.0073535371
## Quantity 0.200794771 1.00000000 0.0086229703 0.066253189 0.0182984399
## Discount -0.028190124 0.00862297 1.0000000000 -0.219487456 0.0004084856
## Profit 0.479064350 0.06625319 -0.2194874564 1.000000000 -0.0046493531
## diff in days -0.007353537 0.01829844 0.0004084856 -0.004649353 1.000000000
```

```
#correlation matrix with statistical significance
cor_result=rcorr(as.matrix(num_data))

cor_result$r
```

```
## Sales Quantity Discount Profit diff_in_days
## Sales 1.000000000 0.20079477 -0.0281901242 0.479064350 -0.0073535371
## Quantity 0.200794771 1.00000000 0.0086229703 0.066253189 0.0182984399
```

```
corrplot(cor_result$r, type = "upper", order = "hclust", tl.col = "black", tl.srt = 45) #display only
```



Discount is negatively correlated with profit, whereas sales is positively correlated with profit. The time between order date and ship date (diff\_in\_days) is not correlated with sales, quantity, discount, or profit.

Since the difference in days between Order date and Ship date has 0 correlation with the other variables, let's drop diff\_in\_days for the K-means clustering analysis.

### **Data Preparation**

```
#make a copy of the original dataset and copy to data1
data1 <- data</pre>
```

drop column Row ID because it is not necessary; it is the row number from the original excel file. The country variable is also not needed because all the values are United states. Customer\_Name and Customer\_ID give redundant information. So we will drop the Customer\_Name column and keep only the Customer\_ID column.

```
data1[,c("Row_ID","i__Row_ID", "Country", "Customer_Name")]<-NULL
```

### head(data1)

```
##
           Order_ID Order_Date Ship_Date
                                                Ship_Mode Customer_ID
                                                                         Segment
## 1 CA-2016-152156 2016-11-08 2016-11-11
                                             Second Class
                                                              CG-12520
                                                                        Consumer
## 2 CA-2016-152156 2016-11-08 2016-11-11
                                             Second Class
                                                              CG-12520
                                                                        Consumer
## 3 CA-2016-138688 2016-06-12 2016-06-16
                                             Second Class
                                                              DV-13045 Corporate
## 4 US-2015-108966 2015-10-11 2015-10-18 Standard Class
                                                              SO-20335
                                                                        Consumer
## 5 US-2015-108966 2015-10-11 2015-10-18 Standard Class
                                                              SO-20335
                                                                        Consumer
  6 CA-2014-115812 2014-06-09 2014-06-14 Standard Class
                                                              BH-11710
                                                                        Consumer
##
                City
                           State Postal_Code Region
                                                          Product_ID
                                                                            Category
## 1
                                       42420
                                              South FUR-B0-10001798
           Henderson
                       Kentucky
                                                                           Furniture
## 2
           Henderson
                       Kentucky
                                       42420
                                              South FUR-CH-10000454
                                                                           Furniture
                                       90036
## 3
         Los Angeles California
                                               West OFF-LA-10000240 Office Supplies
## 4 Fort Lauderdale
                        Florida
                                       33311 South FUR-TA-10000577
                                                                           Furniture
## 5 Fort Lauderdale
                        Florida
                                       33311
                                              South OFF-ST-10000760 Office Supplies
## 6
         Los Angeles California
                                       90032
                                               West FUR-FU-10001487
                                                                           Furniture
##
     Sub Category
                                                                        Product Name
## 1
        Bookcases
                                                  Bush Somerset Collection Bookcase
## 2
           Chairs
                       Hon Deluxe Fabric Upholstered Stacking Chairs, Rounded Back
## 3
           Labels
                         Self-Adhesive Address Labels for Typewriters by Universal
## 4
           Tables
                                      Bretford CR4500 Series Slim Rectangular Table
                                                     Eldon Fold 'N Roll Cart System
## 5
          Storage
     Furnishings Eldon Expressions Wood and Plastic Desk Accessories, Cherry Wood
## 6
##
        Sales Quantity Discount
                                    Profit diff_in_days
## 1 261.9600
                     2
                           0.00
                                   41.9136
                                                       3
## 2 731.9400
                     3
                           0.00
                                  219.5820
                                                       3
     14.6200
                     2
                           0.00
                                                       4
                                    6.8714
                                                       7
## 4 957.5775
                     5
                           0.45 -383.0310
                                                       7
## 5
     22.3680
                     2
                           0.20
                                    2.5164
## 6 48.8600
                     7
                           0.00
                                   14.1694
```

### Model

For this K-means clustering we will use the numeric variables only: which are sales, quantity, discount, profit (columns 15 - 18). K means clustering is affected by the starting assignment points, so we will try with 25 different starting assignments (nstart = 25), and see which ones work the best.

(https://www.datanovia.com/en/blog/k-means-clustering-visualization-in-r-step-by-step-guide/)

```
## [8252] 1 2 2 2 2 2 2 2 2 2 2 2 2 1 2 2 2 1 2 2 2 2 2 2 2 2 1 1 2 2 2 2 2 2 1 2 2
## [9991] 2 2 2 2
```

summary(results\_kmeans)

##

Length Class Mode

```
## cluster
                9994
                       -none- numeric
## centers
                  12
                      -none- numeric
## totss
                       -none- numeric
                   3
## withinss
                       -none- numeric
## tot.withinss
                   1
                       -none- numeric
## betweenss
                   1
                       -none- numeric
## size
                   3
                       -none- numeric
## iter
                   1
                       -none- numeric
## ifault
                       -none- numeric
```

### results\_kmeans

```
## K-means clustering with 3 clusters of sizes 1136, 8831, 27
##
## Cluster means:
##
 Sales
  Quantity
   Discount
     Profit
## 1 -0.05714414 0.045908572 2.3730228 -0.5928228
## 2 -0.02922217 -0.007823215 -0.3039892 0.0425665
## 3 11.96210167  0.627210176  -0.4157497 11.0200717
##
## Clustering vector:
##
##
##
##
##
##
##
##
##
##
##
##
##
##
##
##
##
##
##
##
##
##
##
```

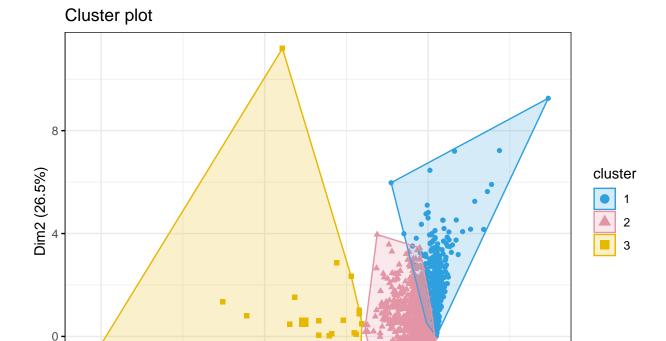
```
## [9991] 2 2 2 2
##
## Within cluster sum of squares by cluster:
  5207.851 16648.669 3315.203
##
 (between_SS / total_SS = 37.0 %)
##
## Available components:
##
## [1] "cluster"
      "centers"
           "totss"
               "withinss"
                    "tot.withinss"
## [6] "betweenss"
      "size"
           "iter"
               "ifault"
# https://towardsdatascience.com/clustering-analysis-in-r-using-k-means-73eca4fb7967
#cluster means are the centroid vectors
#clustering vector is the group that the observation is placed into
#percentage indicates compactness of the clustering or how similar observations are within the same gro
```

The results of this clustering indicate that the within cluster sum of squares by cluster is 37.0 % which means that the observations within a given group are not very similar to each other.

# Plot K-means

The factoextra package contains a function called fviz\_cluster() which can be used to visualize kmeans clusters. The input required is the original dataset, and the kmeans results. These are used to produce plots which show points that represent observations.

```
ggtheme = theme_bw()
)
```



Reduce dimensions using Principal Component Analysis.

<del>-</del>20

-40

```
results_pca <- prcomp(data1[,(15:19)], scale=TRUE)

#Coordinates of individual observations
indiv_coordinates <- as.data.frame(get_pca_ind(results_pca)$coord)

#Add clusters obtained through the Kmeans algorithm
indiv_coordinates$cluster <- factor(results_kmeans$cluster)

#Add region from the dataset
indiv_coordinates$Region <- data1$Region

#look at the first few rows of individual coordinates
head(indiv_coordinates)</pre>
```

Dim1 (39.7%)

Ö

```
##
                      Dim.2
                                Dim.3
                                                          Dim.5 cluster Region
          Dim.1
                                              Dim.4
## 1 0.04520418 -1.13666870 0.3612739 0.0807637427
                                                     0.30768311
                                                                      2 South
## 2 1.15718390 -0.84275631 0.4252513 0.1851981706 0.21856052
                                                                        South
## 3 -0.31072375 -1.08484635 -0.2274155 -0.0007838382 0.15040037
                                                                         West
## 4 -0.63620673 2.24187253 -1.2663321 0.5805617103 1.55922466
                                                                     1 South
## 5 -0.58444157 -0.08714015 -1.6947210 0.7880697439 -0.05040709
                                                                     2 South
## 6 0.39080002 0.44622564 -0.7312599 -1.4739348647 -0.19665673
                                                                         West
```

Percentage of variance explained by dimensions.

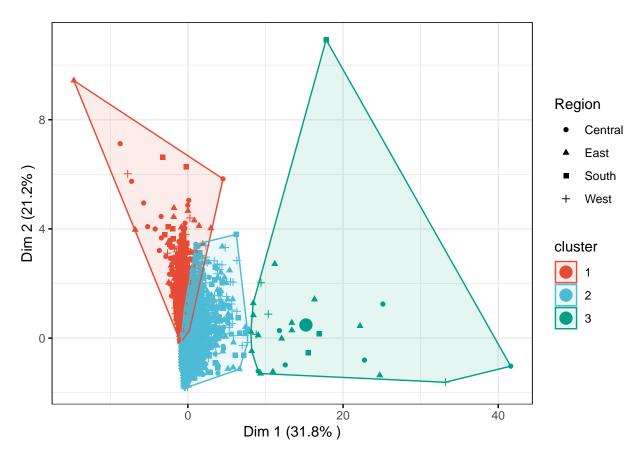
```
eigenvalue <- round(get_eigenvalue(results_pca), 1)
variance.percent <- eigenvalue$variance.percent
head(eigenvalue)</pre>
```

```
##
         eigenvalue variance.percent cumulative.variance.percent
## Dim.1
                1.6
                                 31.8
                                                              31.8
                                 21.2
## Dim.2
                1.1
                                                              53.0
## Dim.3
                1.0
                                 20.0
                                                              73.0
## Dim.4
                0.9
                                 17.6
                                                              90.6
## Dim.5
                                  9.4
                                                             100.0
                0.5
```

Variance of a group indicates how different members of a group are. Higher variance means greater dissimilarity within a group.

```
#To visualize the k-means clusters:

ggscatter(
  indiv_coordinates, x = "Dim.1", y = "Dim.2",
  color = "cluster", palette = "npg", ellipse = TRUE, ellipse.type = "convex", #adding the concentral
  shape = "Region", size = 1.5, legend = "right", ggtheme = theme_bw(),
  xlab = paste0("Dim 1 (", variance.percent[1], "%)"),
  ylab = paste0("Dim 2 (", variance.percent[2], "%)")
) +
  stat_mean(aes(color = cluster), size = 4) #stat_mean is used for adding the cluster centroid
```



The clustering plot shows that the groups are very close together, and overlap slightly. The clusters could be further apart with some tuning by changing the number of clusters (k).

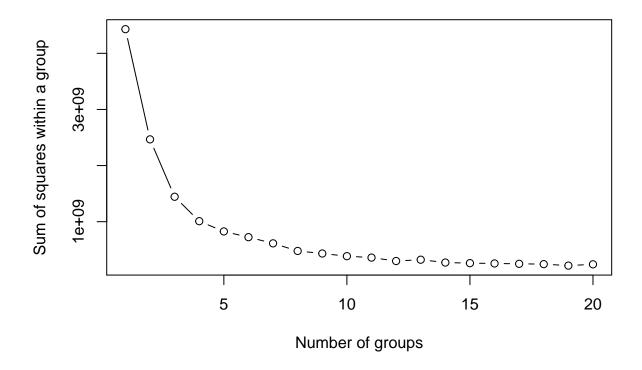
### **Evaluation**

(https://towardsdatascience.com/clustering-analysis-in-r-using-k-means-73eca4fb7967)

The within sum of squares (Withinss) is a value that represents the level of dissimilarity within a group. The higher the withinss, the greater the dissimilarity within the group.

(Foncesca, 2019)

```
## Warning: did not converge in 10 iterations
```



From the Within sum of squares plot, the optimal number of clusters is around 5. when the number of groups (k) initially increases 1 to 6, the error measures (sum of squares within a group) starts to decrease. When the number of groups is 7 or 8, the error measure starts to flatten.

The main purpose is to find a number of initial groups which achieves some fair amount of compactness (or similarity) between observations within a group. When k is too high, each cluster starts to represent individual points, whereas when k is too low, the observations may not be in the right cluster.

We can try re-running the k-means model with the number of groups, k=4

##

```
set.seed(123)
clustering_results_4 <- kmeans(scale(data1[,(15:18)]), centers = 4, nstart = 25)
clustering results 4
## K-means clustering with 4 clusters of sizes 27, 1044, 2838, 6085
##
## Cluster means:
##
         Sales
                 Quantity
                          Discount
## 1 11.9621017
               0.62721018 -0.4157497 11.02007172
  2 -0.1222964
              0.02095769
                         2.4788189 -0.58466503
## 3 0.2856822
              1.19688927 -0.2981228 0.17558930
  4 -0.1653353 -0.56459922 -0.2844025 -0.03048054
##
## Clustering vector:
##
     [38] 3 4 4 4 4 4 4 4 4 4 4 4 4 3 3 4 4 3 3 3 3 3 3 4 4 4 4 4 3 3 4 4 3 3 4 4 4 4 3 3
##
```

```
## [8733] 4 4 3 4 4 2 4 4 3 4 4 4 4 4 4 4 4 4 3 4 4 4 3 2 2 4 3 4 4 4 3 4 4 4 2 2 2 4
## [8770] 4 4 3 4 3 4 4 3 3 4 4 4 3 4 4 4 3 4 4 4 4 4 3 3 4 4 4 4 3 3 4 4 4 3 3 4 4 4 4 4 4 4
## [9029] 4 4 4 4 4 4 4 2 4 4 4 4 1 2 4 4 4 4 4 3 3 3 3 4 4 4 4 4 4 4 4 3 3 4 3 4 4 3 3 2
## [9103] 4 2 2 4 2 4 3 4 3 4 4 4 3 2 4 4 4 3 3 3 4 2 2 2 4 4 2 2 4 4 4 3 4 4 4 2 3
## [9214] 4 3 3 2 4 3 2 4 3 4 2 2 2 4 3 4 4 3 2 3 4 3 3 4 4 4 3 2 4 3 4 4 3 4 4 3 4 4 3
## [9362] 3 3 4 4 4 4 3 3 4 4 4 4 3 3 3 3 4 4 4 4 4 4 4 4 4 3 4 3 4 4 4 4 4 3 4 3 3 3 4 4 3 4
## [9547] 4 4 2 4 4 3 4 3 3 4 3 2 3 3 4 4 2 4 4 3 2 4 4 2 4 4 3 4 4 2 3 4 4 3 3 3
## [9880] 3 4 4 3 4 4 4 4 4 4 4 4 4 3 3 4 3 3 4 4 4 4 4 4 4 4 4 2 4 3 4 4 3 4 4 3 4 4 4 3
## [9991] 4 4 4 4
##
```

## Within cluster sum of squares by cluster:

```
## [1] 3315.203 4284.643 7571.058 3614.724
## (between_SS / total_SS = 53.0 %)
##
## Available components:
##
## [1] "cluster" "centers" "totss" "withinss" "tot.withinss"
## [6] "betweenss" "size" "iter" "ifault"
```

The within cluster sum of squares by cluster value is now 53.0%. This represents the compactness of the clustering, or how similar observations are to other observations within the same group.

We can try re-running the k-means model with the number of groups, k = 7

```
set.seed(123)
clustering_results_7 <- kmeans(scale(data1[,(15:18)]), centers = 7, nstart = 25)</pre>
clustering_results_7
## K-means clustering with 7 clusters of sizes 3317, 200, 12, 914, 2574, 9, 2968
##
## Cluster means:
##
            Sales
                       Quantity
                                   Discount
                                                     Profit
## 1 -0.17099202 -0.55401551 -0.7476979
                                                0.01999970
## 2 3.70037129 1.13047292 -0.4332859
                                                2.62598142
## 3 8.36049519
                    1.14320638 2.0689103 -11.91366497
## 4 -0.21824012 0.05129912 2.6451425
                                               -0.46197164
## 5 0.04223395 1.22666437 -0.2946436
                                                0.03793345
## 6 16.87915324 0.84359568 -0.6489669
                                               19.79865024
## 7 -0.11265862 -0.54381990 0.2993725
                                               -0.10180518
##
## Clustering vector:
##
        \begin{smallmatrix} 1 \end{smallmatrix} 1 \ 1 \ 1 \ 1 \ 4 \ 7 \ 5 \ 1 \ 5 \ 7 \ 5 \ 5 \ 7 \ 7 \ 4 \ 4 \ 5 \ 1 \ 1 \ 7 \ 7 \ 5 \ 5 \ 7 \ 1 \ 7 \ 1 \ 3 \ 4 \ 7 \ 7 \ 5 \ 4 \ 7 \ 7 \ 5 \ 4 
      ##
##
     [75] \ 7 \ 4 \ 4 \ 7 \ 4 \ 1 \ 1 \ 5 \ 1 \ 5 \ 7 \ 1 \ 1 \ 1 \ 5 \ 1 \ 7 \ 1 \ 1 \ 1 \ 4 \ 5 \ 7 \ 5 \ 5 \ 7 \ 4 \ 1 \ 5 \ 7 \ 4 \ 5 \ 7 \ 7 \ 5 \ 1
    [112] 1 5 5 7 7 5 1 4 1 1 5 5 5 7 4 4 5 7 1 7 7 5 1 1 1 1 7 5 5 7 1 1 1 1 5 7 5
##
    [149] \ 1 \ 2 \ 5 \ 5 \ 7 \ 5 \ 1 \ 5 \ 1 \ 7 \ 1 \ 5 \ 1 \ 7 \ 5 \ 7 \ 7 \ 3 \ 7 \ 5 \ 5 \ 4 \ 7 \ 1 \ 1 \ 5 \ 4 \ 7 \ 4 \ 5 \ 5 \ 1 \ 7 \ 5 \ 1 \ 5 \ 1
##
##
    [186] 1 5 7 7 7 5 5 7 5 5 7 7 1 4 7 7 4 7 4 5 7 1 1 1 5 1 7 7 4 7 4 7 5 7 7 5 4
    [223] 7 4 7 1 5 1 7 5 4 7 4 4 7 7 1 5 4 4 7 4 4 7 2 1 5 2 1 7 1 2 5 1 4 7 5 5 1
##
    [260] 1 5 4 7 5 7 1 7 5 7 7 1 5 1 7 1 1 7 7 5 7 4 5 5 5 5 5 4 7 5 5 5 7 4 5 7 7
    [297] \ 4 \ 7 \ 5 \ 5 \ 1 \ 1 \ 1 \ 4 \ 1 \ 1 \ 5 \ 1 \ 7 \ 7 \ 7 \ 1 \ 7 \ 7 \ 7 \ 1 \ 2 \ 5 \ 1 \ 1 \ 1 \ 5 \ 1 \ 7 \ 7 \ 7 \ 5 \ 4 \ 7 \ 4
##
##
    ##
   [371] 5 7 4 5 7 5 7 5 4 7 4 5 7 1 7 7 4 7 7 7 1 7 2 4 7 1 1 1 7 7 7 7 7 7 5 1 5 7
    [408] \ 5\ 5\ 1\ 5\ 1\ 5\ 7\ 5\ 5\ 1\ 1\ 1\ 5\ 1\ 1\ 1\ 5\ 1\ 5\ 4\ 5\ 5\ 7\ 5\ 4\ 5\ 7\ 5\ 5\ 1\ 1\ 1\ 1\ 1
    ##
##
    [482] \ 1 \ 1 \ 1 \ 1 \ 5 \ 5 \ 2 \ 7 \ 7 \ 5 \ 1 \ 1 \ 7 \ 1 \ 5 \ 1 \ 7 \ 5 \ 4 \ 4 \ 7 \ 4 \ 7 \ 5 \ 1 \ 1 \ 1 \ 6 \ 1 \ 1 \ 7 \ 1 \ 1 \ 2 \ 1 \ 1
##
    [519] 1 7 7 4 1 7 1 5 7 7 7 7 5 7 1 5 5 7 4 7 1 1 5 5 5 7 7 4 5 1 5 4 5 7 7 7 1
    [556] \ 7 \ 1 \ 7 \ 5 \ 1 \ 5 \ 5 \ 7 \ 1 \ 1 \ 7 \ 5 \ 5 \ 1 \ 7 \ 5 \ 1 \ 5 \ 1 \ 1 \ 5 \ 1 \ 1 \ 7 \ 4 \ 4 \ 1 \ 7 \ 7 \ 4 \ 7 \ 1 \ 1 \ 7 \ 7 \ 4 \ 7
##
##
    [593] 7717557777744774577775541151141515515
    [630] \ 1 \ 7 \ 5 \ 7 \ 7 \ 1 \ 4 \ 7 \ 5 \ 7 \ 1 \ 1 \ 5 \ 1 \ 1 \ 1 \ 5 \ 7 \ 1 \ 7 \ 1 \ 5 \ 1 \ 5 \ 7 \ 4 \ 7 \ 7 \ 4 \ 5 \ 5 \ 5 \ 4 \ 5 \ 1 \ 7
##
    [667] \ 7 \ 7 \ 5 \ 4 \ 7 \ 5 \ 5 \ 1 \ 5 \ 1 \ 4 \ 7 \ 4 \ 4 \ 7 \ 1 \ 1 \ 3 \ 7 \ 1 \ 1 \ 1 \ 1 \ 1 \ 1 \ 5 \ 5 \ 7 \ 5 \ 7 \ 7 \ 4 \ 4 \ 7 \ 1
    ##
    [741] 7 4 5 7 7 7 5 1 7 1 5 5 1 7 1 4 1 1 7 1 5 1 7 1 1 5 1 5 7 1 1 1 5 1 5 5
##
##
    [778] \ 7 \ 1 \ 1 \ 1 \ 4 \ 7 \ 7 \ 7 \ 1 \ 7 \ 7 \ 1 \ 1 \ 5 \ 1 \ 5 \ 5 \ 1 \ 5 \ 5 \ 1 \ 7 \ 5 \ 1 \ 1 \ 7 \ 2 \ 1 \ 1 \ 5 \ 1 \ 1 \ 1
    [815] \ 2\ 5\ 7\ 5\ 4\ 5\ 4\ 7\ 1\ 1\ 1\ 1\ 5\ 1\ 1\ 1\ 1\ 1\ 7\ 7\ 7\ 4\ 7\ 7\ 1\ 1\ 7\ 7\ 7\ 1\ 1\ 5\ 1\ 7\ 5\ 1
```

[852] 5 5 1 1 1 5 1 1 1 1 1 5 7 7 1 1 7 4 4 7 1 1 5 5 1 1 1 4 7 5 1 1 5 1 7 5 5

##

```
[889] 5 5 7 1 1 5 5 7 7 5 5 4 5 4 7 1 1 1 7 5 7 1 2 1 5 1 5 7 7 4 7 7 1 1 1 5 7
   [926] \ 7\ 5\ 1\ 1\ 7\ 7\ 7\ 5\ 4\ 7\ 7\ 1\ 1\ 1\ 1\ 7\ 7\ 5\ 1\ 5\ 4\ 5\ 5\ 5\ 7\ 5\ 7\ 7\ 7\ 5\ 7\ 4\ 7\ 1\ 7\ 1
##
   [963] 1 7 1 1 1 1 1 1 5 7 4 7 5 7 4 2 4 1 5 7 7 5 7 7 7 4 5 1 5 7 1 5 2 1 1 1 5
## [1000] 5 1 2 5 1 1 7 7 4 5 7 1 1 1 2 1 5 1 1 5 5 1 1 2 7 7 7 7 7 5 1 1 5 5 7 4 4
## [1037] 1 7 5 7 1 1 7 5 4 5 5 7 5 1 7 7 5 1 1 7 7 7 7 7 7 7 7 5 7 7 7 4 7 7 4 5 7 7
## [1074] 4 4 7 5 1 1 7 1 5 5 1 1 2 1 7 1 7 7 5 7 5 1 5 4 5 1 1 1 1 4 4 1 5 4 4 5 7
## [1148] 1 1 1 1 5 1 5 1 2 1 1 7 1 1 1 1 7 5 1 5 1 1 1 1 1 5 4 5 1 1 1 1 7 5 1 7 5
## [1185] 5 5 1 1 1 5 7 1 1 7 7 7 7 5 7 4 7 4 7 7 7 5 1 7 5 5 7 7 7 4 4 7 7 7 7 5 5
## [1259] 7 5 7 4 7 1 5 1 1 5 7 1 7 4 4 7 7 7 7 7 1 1 1 7 4 7 4 5 1 1 7 1 7 5 1 5 1
## [1296] 5 7 7 5 1 5 1 1 1 1 1 5 5 7 1 1 7 1 1 7 5 5 5 7 7 4 4 7 5 7 1 4 5 1 1 1 7 7
## [1333] 4 4 5 1 1 1 1 5 4 4 4 1 5 1 5 5 5 5 7 1 5 1 7 7 5 1 4 7 5 7 7 5 4 5 7 7 5
## [1370] 4 7 1 1 7 5 1 5 5 5 5 7 4 1 5 5 1 1 1 1 1 1 1 5 7 4 4 7 1 1 1 1 1 7 5 7 7
## [1407] 7 1 1 4 7 1 1 5 1 1 5 7 4 7 4 7 4 5 5 7 7 5 5 5 5 5 1 2 5 5 7 7 5 5 1 7 1
## [1444] 7 2 7 5 7 4 7 4 5 1 5 2 1 1 5 7 1 7 7 7 5 1 7 7 7 1 1 1 7 1 4 5 7 5 5 7 1
## [1629] 5 4 7 7 7 7 7 7 7 5 7 1 1 1 1 7 1 2 1 7 7 4 7 1 1 5 1 1 7 1 5 7 7 7 7 7 4 5 7
## [1666] 7 7 7 7 7 7 1 7 1 4 4 1 1 7 7 7 4 1 1 7 4 4 7 7 5 5 5 5 7 1 1 1 1 5 5 7 2
## [1740] 5 5 7 7 7 5 1 4 7 7 7 7 1 5 1 1 5 5 4 7 7 7 5 7 1 1 5 1 4 4 7 4 7 5 1 5 1
## [1777] 1 5 5 1 1 1 1 1 5 5 1 1 1 1 5 5 2 1 7 1 5 1 5 5 7 7 1 4 3 7 2 5 5 5 4 7 5 5
## [1814] 7 7 1 1 1 4 4 4 7 7 7 4 7 7 7 5 7 2 1 7 7 5 5 5 7 7 7 7 1 7 4 1 5 1 7 7 1
## [1851] 5 5 5 1 5 1 5 4 1 1 7 1 7 4 7 7 5 5 7 1 5 7 7 1 7 7 7 7 7 4 1 5 1 5 7 1 7
## [1888] 4 7 7 1 1 7 7 1 7 2 7 4 7 5 1 7 7 7 5 4 4 1 1 1 5 7 5 7 7 7 5 1 5 1 7 7 5 7
## [1925] 1 1 7 7 1 1 7 7 1 5 5 5 1 4 5 4 7 7 7 4 7 7 1 1 7 1 5 1 5 5 1 1 1 1 1 1 5
## [1962] 5 1 5 1 1 1 5 1 1 7 5 4 7 7 1 2 5 7 7 7 1 1 5 1 1 7 1 5 1 1 4 5 7 7 4 4 1
## [1999] 1 5 5 5 1 5 4 5 5 7 7 7 7 7 7 1 1 1 7 1 1 5 2 1 1 1 5 1 7 7 1 5 1 1 7 7 1
## [2036] 1 1 1 7 4 4 5 5 5 7 1 1 1 5 4 4 5 5 5 1 5 2 1 1 1 5 7 5 7 1 7 5 1 1 7 4 5
## [2073] 7 1 5 1 5 5 5 5 7 7 5 7 1 5 5 1 1 1 1 7 5 1 5 5 1 1 7 5 7 7 7 1 1 7 4 1 7
## [2110] 1 5 7 1 7 1 7 7 5 1 5 4 7 1 5 5 5 7 7 7 1 5 5 1 4 5 7 4 4 1 5 1 7 5 7 5 7
## [2147] 4 5 1 5 7 5 1 7 4 5 1 1 7 4 7 1 1 7 7 4 7 4 7 7 7 7 7 7 4 7 1 5 5 5 5 1 1 2
## [2184] 5 5 7 1 5 1 2 1 7 1 1 1 5 7 1 7 7 5 5 7 1 7 7 4 7 1 5 5 5 1 7 5 1 7 7 7 1
## [2221] 7 7 5 7 7 4 7 7 7 7 1 5 5 7 5 1 5 5 4 7 5 7 1 1 1 1 7 1 5 5 5 1 1 1 1 1 1
## [2258] 5 1 1 1 1 1 1 1 1 7 1 1 7 5 5 5 7 7 7 5 1 1 7 5 1 5 7 1 5 7 1 5 5 1 1 5 1 1
## [2332] 5 5 1 5 1 5 1 1 7 5 5 1 7 5 5 4 5 5 7 1 1 5 5 5 1 7 5 4 7 5 1 4 1 1 1 7 1
## [2406] 5 5 5 7 5 7 5 1 7 5 1 5 7 2 1 1 7 7 5 4 7 5 4 4 4 5 1 5 7 1 1 5 4 7 7 5 7
## [2480] 7 7 5 7 5 5 1 1 1 1 1 1 5 2 7 5 5 1 1 7 7 7 4 4 1 1 2 5 4 7 5 1 5 4 7 7 7
## [2517] 5 1 5 5 1 1 1 1 5 7 4 4 7 7 5 1 5 1 5 1 7 7 7 5 7 5 7 1 1 1 1 7 7 2 5 7 5 1
## [2554] 1 5 1 4 7 5 5 1 7 7 1 1 1 5 2 5 7 7 7 5 5 1 5 1 1 1 5 1 1 5 1 1 7 5 1 1 7
## [2591] 4 4 1 5 1 1 1 1 1 1 5 1 1 7 1 4 1 7 7 5 7 7 4 7 7 5 1 5 1 7 1 5 1 6 2 5 1
## [2628] 7 4 4 1 1 7 7 7 4 1 1 7 1 1 1 5 7 5 1 1 1 5 1 5 7 7 1 1 7 1 5 1 7 2 1 1 7
## [2665] 7 4 7 7 7 1 1 5 1 5 1 5 7 4 4 5 5 1 1 1 5 7 4 4 5 7 4 7 7 5 7 7 4 3 7 7 7
## [2702] 7 7 1 5 1 1 1 5 2 4 7 5 7 5 1 1 1 1 4 7 5 5 5 7 7 7 4 5 1 1 1 1 1 1 1 5 1
## [2739] 1 1 1 5 1 5 7 7 7 7 4 7 5 7 1 1 5 2 1 1 1 1 5 1 5 4 4 7 7 2 4 5 7 5 7 1 7
## [2776] 1 5 7 5 7 1 5 1 7 7 7 1 4 5 4 1 1 1 5 7 5 7 1 1 5 5 7 1 1 5 7 1 1 5 7 5 5
## [2813] 1 1 1 5 1 1 5 5 7 7 7 7 5 7 4 7 5 5 7 5 7 5 1 1 1 5 4 1 1 5 7 7 5 1 4 5 2
## [2850] 4 7 1 1 1 1 5 7 7 1 1 2 5 7 1 5 1 1 5 7 7 7 7 1 5 1 7 5 1 1 1 5 1 4 1 7 5
```

```
## [2924] 1 1 1 1 1 4 1 7 7 7 7 7 1 5 5 5 7 1 2 1 5 1 5 1 5 5 7 5 7 7 7 7 1 1 4 7 1
## [2961] 1 5 7 7 5 1 4 7 7 5 4 7 7 5 7 7 5 1 5 7 7 7 7 7 4 1 1 5 1 7 5 5 1 1 7 5 7
## [2998] 7 7 4 7 4 5 7 1 1 5 7 1 5 4 3 4 4 7 7 5 4 5 1 1 1 5 7 5 1 7 7 5 7 1 1 7 7
## [3035] 5 5 1 5 1 7 5 1 1 2 7 5 1 5 1 1 1 1 1 1 1 1 2 1 5 1 7 7 5 1 1 5 5 5 5 5 7 2
## [3072] 1 5 1 7 7 4 7 5 1 7 1 1 7 4 7 4 4 5 5 5 5 1 7 7 5 4 7 7 7 7 5 1 4 5 7 5 1
## [3109] 1 1 5 1 7 5 5 5 5 1 5 7 7 7 4 4 7 7 7 1 1 7 5 7 1 7 4 7 7 7 7 4 1 4 5 5 5
## [3146] 7 7 5 5 1 2 3 7 5 1 7 7 1 5 7 4 1 5 1 5 7 1 5 5 5 5 5 4 5 7 5 4 7 1 1 1 5 1
## [3183] 1 2 1 4 7 5 1 7 7 1 4 4 5 1 1 4 7 5 7 7 5 1 1 5 5 5 5 5 7 4 5 7 7 5 1 7 1
## [3220] 5 1 1 7 7 7 5 7 7 7 1 5 4 4 7 7 7 1 7 7 5 4 5 4 5 7 1 5 1 7 7 5 7 7 7 7 7
## [3257] 7 7 4 4 7 1 5 1 7 7 7 1 1 1 1 5 5 2 1 5 5 1 1 5 2 7 7 1 4 1 1 7 5 5 1 1 1
## [3294] 1 4 2 4 7 2 4 7 5 7 4 7 7 7 4 7 5 5 5 5 1 1 7 4 5 1 5 7 5 1 1 4 4 7 5 5 7
## [3368] 7 1 7 5 5 1 1 5 1 1 5 4 7 5 7 1 5 2 7 7 7 4 7 7 7 7 4 5 1 1 1 1 5 1 1 4 4
## [3405] 1 7 7 4 7 7 5 7 1 1 7 1 1 5 7 1 1 5 5 7 7 1 2 7 7 5 1 1 5 7 5 4 4 4 7 5 7
## [3442] 5 1 2 7 5 1 1 7 5 1 1 7 1 7 7 7 7 4 4 7 7 5 7 7 1 1 5 5 7 1 7 7 1 1 5 5 5
## [3479] 1 5 5 1 5 1 1 1 1 7 7 7 5 5 7 7 1 5 1 7 5 5 1 1 5 7 7 1 7 4 7 7 7 7 5 5 7 7
## [3516] 1 5 7 7 4 7 5 1 5 5 1 1 5 7 7 5 5 7 1 5 1 1 1 1 5 5 7 7 7 5 4 5 1 1 5 7 5
## [3553] 5 4 1 5 5 7 7 1 7 4 4 7 1 1 2 4 1 2 2 7 7 1 5 7 7 1 1 5 2 1 1 1 1 1 1 1 7 1
## [3627] 1 1 2 7 5 1 5 7 5 7 7 7 5 7 5 4 7 1 1 5 1 1 4 5 5 5 1 1 5 7 7 7 7 7 5 1 5
## [3664] 1 5 4 7 5 5 5 7 7 7 4 7 7 1 1 1 1 1 1 1 5 5 5 1 4 5 5 4 7 5 7 5 5 1 1 5 1
## [3701] 1 7 5 1 1 5 7 4 4 5 7 5 1 1 1 5 5 1 5 5 1 1 5 4 7 7 1 1 1 1 5 1 5 1 1 1 1
## [3738] 1 1 1 1 1 5 7 1 5 1 1 1 7 5 7 1 7 5 5 7 5 1 1 1 5 1 1 4 7 7 5 5 7 5 7 1 7
## [3775] 7 7 5 5 7 1 1 7 7 7 7 4 5 7 7 5 5 4 5 1 1 7 1 7 7 5 7 7 7 7 7 7 5 4 4 5 7
## [3812] 7 1 5 5 7 1 1 5 7 7 1 5 5 7 5 5 2 1 1 1 1 1 7 7 7 7 7 7 7 5 1 7 7 1 1 5 5 1
## [3849] 1 5 5 1 4 1 5 4 5 7 5 4 7 7 7 1 4 1 5 5 7 5 1 1 7 7 4 7 7 4 7 7 7 7 5 7 7
## [3886] 4 1 1 1 1 1 2 5 5 4 5 7 7 7 7 7 7 5 5 7 4 7 7 7 7 7 4 7 1 1 4 7 7 4 7 1 1 7
## [3923] 5 5 1 1 1 5 7 7 5 7 7 5 5 7 7 7 7 2 1 7 5 5 1 7 7 7 7 7 5 7 7 1 1 1 1 1 4 7
## [3960] 7 5 5 1 1 7 7 1 4 4 5 5 1 1 5 5 5 1 1 7 7 1 1 7 2 1 7 2 5 5 1 7 7 7 5 1 5
## [4071] 4 7 1 5 7 7 1 1 1 7 5 4 4 7 4 7 7 7 5 1 1 7 7 2 1 1 5 5 6 1 4 4 5 1 5 7 4
## [4108] 1 7 1 1 1 1 7 7 5 7 1 5 5 7 5 1 7 1 7 5 5 4 1 4 7 7 5 1 5 5 5 4 7 5 1 5 5
## [4182] 1 1 1 1 5 4 4 7 1 6 5 1 7 1 5 1 7 4 1 7 5 5 5 1 1 1 1 2 7 7 1 1 1 1 1 1 5
## [4219] 2 1 5 5 1 1 4 4 1 7 1 4 4 7 5 4 7 5 4 5 1 1 1 7 1 1 1 1 5 5 5 1 5 5 7 7 7
## [4256] 4 4 4 7 1 5 1 1 4 5 4 5 4 5 1 5 1 5 1 7 4 1 6 1 5 1 1 1 1 7 7 7 7 4 7 1 5
## [4293] 1 7 4 1 1 2 7 5 7 5 7 5 1 5 7 7 5 7 7 1 7 1 1 7 1 1 5 7 1 4 7 4 7 7 1 5 5
## [4330] 7 5 1 5 5 7 7 7 1 1 7 1 1 5 5 1 1 1 1 4 5 7 7 7 1 4 4 7 7 4 7 1 1 1 1 1 1 1
## [4367] 5 7 5 5 1 5 1 5 5 5 1 5 1 1 1 7 4 7 7 5 1 7 5 7 1 1 7 5 7 4 7 7 7 5 1 5 7
## [4404] 1 7 7 1 1 1 1 5 1 1 1 1 7 5 5 7 4 5 7 1 1 5 5 1 1 2 1 1 1 5 1 1 4 5 4 7 7
## [4478] 1 1 5 5 1 5 1 7 1 1 4 5 4 7 7 7 5 5 5 5 4 1 1 5 5 5 5 1 1 7 1 1 7 1 1 4 7
## [4515] 1 1 5 5 1 5 1 1 7 5 5 1 5 1 7 4 1 1 7 1 1 1 7 1 1 1 5 4 1 5 7 5 4 7 4 7 1
## [4552] 7 7 1 1 1 1 4 7 7 1 7 5 7 5 7 4 7 4 1 1 1 7 5 1 5 1 7 1 5 5 7 5 1 5 5 5 1
## [4589] 1 5 1 7 7 1 7 1 5 1 5 1 5 1 7 7 7 5 1 1 7 5 5 1 7 5 4 5 7 1 1 2 7 5 7 7 7
## [4663] 7 7 4 7 1 1 4 5 5 1 1 1 1 7 7 1 5 1 7 5 5 7 4 4 5 5 4 7 7 7 5 7 1 1 1 5 1
## [4700] 7 7 7 4 4 5 1 1 5 1 1 1 4 5 4 7 7 1 1 5 5 1 4 1 1 4 7 1 7 5 7 5 1 1 1 1 1
## [4737] 1 1 1 7 1 1 1 1 1 1 1 1 1 1 5 7 7 7 7 4 7 4 4 1 7 7 7 7 7 1 1 1 5 7 7 7 1 5 4
## [4774] 5 7 1 1 1 5 1 7 4 5 5 5 4 7 7 5 5 4 7 7 1 1 1 4 5 7 5 7 4 4 7 5 1 5 1 7 1
## [4811] 1 7 7 7 1 7 7 7 7 1 4 5 4 4 7 7 7 1 1 7 5 1 1 1 5 1 7 4 5 7 1 4 5 1 7 7 5
## [4848] 5 7 1 5 5 5 1 1 1 7 1 5 5 2 5 7 7 1 2 5 1 1 7 7 5 7 4 7 4 5 7 7 4 1 1 2 1 5
```

```
## [4885] 1 5 1 5 1 5 1 1 1 1 1 1 1 7 7 4 1 1 5 1 7 1 5 4 7 7 5 7 4 4 1 7 5 7 1 1 4
## [4959] 5 7 1 7 1 1 4 7 1 1 1 1 7 7 7 7 1 5 5 7 5 7 4 7 1 1 5 5 7 5 4 7 7 3 5 5 5
## [4996] 7 7 4 7 5 1 5 1 1 1 5 2 1 4 5 4 4 4 5 5 5 7 1 1 1 1 1 1 1 1 1 1 1 7 1 1 5
## [5033] 1 1 5 1 1 5 7 1 5 1 1 5 5 1 7 5 4 7 1 7 1 5 7 5 5 5 4 1 5 1 1 1 1 7 5 5 1
## [5070] 1 1 5 1 1 7 5 5 7 1 7 1 7 7 7 1 4 5 7 4 7 7 7 1 4 4 7 4 4 7 7 7 7 7 5 7 7
## [5107] 1 5 5 5 5 7 4 7 7 7 1 1 5 5 5 1 1 1 1 5 5 2 1 1 1 5 5 7 5 7 5 7 1 1 5 1 5 5
## [5144] 7 1 1 1 7 7 1 1 5 7 1 1 1 7 7 4 7 7 5 1 7 1 1 1 1 1 1 2 1 1 7 4 7 7 7 1 7
## [5181] 1 5 1 1 5 5 5 1 2 5 7 1 5 5 1 1 1 5 2 7 7 5 5 1 1 1 1 1 5 1 1 1 1 1 5 5 1
## [5292] 5 7 7 5 7 1 5 1 1 2 2 5 7 4 5 1 1 7 5 3 7 7 1 1 5 7 7 7 4 4 4 7 1 5 5 1 1
## [5403] 7 7 5 1 1 1 7 7 1 5 1 5 4 7 7 1 5 7 7 7 7 5 1 2 5 5 1 7 7 7 7 4 7 2 1 7 4 7
## [5440] 7 7 7 1 5 1 7 5 5 1 5 1 1 5 5 7 7 5 1 5 5 7 1 4 7 5 4 7 7 7 1 5 5 7 4 1 1
## [5477] 5 1 1 4 7 1 1 1 7 7 4 7 7 7 7 7 1 1 7 4 7 1 1 1 5 1 7 7 7 7 1 1 5 7 7 7 7
## [5514] 1 1 1 2 5 7 7 5 7 5 1 5 1 1 1 1 5 7 2 5 1 1 1 1 1 1 1 1 7 4 7 4 4 4 5 1 7 1 5
## [5551] 7 5 7 5 7 1 1 1 1 5 1 1 2 1 7 7 5 7 7 7 7 7 4 7 7 7 5 1 1 5 5 1 5 1 1 1 1
## [5662] 7 5 7 4 7 7 1 5 1 1 7 7 7 1 5 1 4 5 1 1 1 1 5 5 7 5 5 5 7 5 5 1 5 1 7 4 4
## [5736] 7 1 1 7 1 1 7 1 1 5 5 7 7 7 5 4 1 5 1 1 4 2 7 4 1 7 1 1 5 5 7 5 5 1 1 1 1
## [5773] 1 7 5 7 7 1 1 7 7 1 1 1 4 5 4 5 5 4 5 7 1 1 7 2 1 1 1 7 7 7 7 7 7 7 7 7 7 7
## [5847] 7 1 1 1 1 1 1 1 5 7 4 4 7 1 7 1 1 1 1 1 5 7 4 7 5 1 1 5 1 7 5 1 5 1 7 5 5
## [5884] 1 2 7 1 1 1 1 7 7 7 1 1 1 5 1 1 1 5 1 1 1 5 1 1 7 1 1 7 1 1 5 5 1 1 2 5 1
## [5995] 7 7 7 7 7 1 1 7 4 4 1 1 1 1 1 1 2 1 7 5 2 1 5 1 1 5 1 1 7 1 5 7 5 5 7 1 7
## [6032] 1 5 1 4 4 1 5 1 1 7 7 7 7 5 7 1 4 4 5 5 5 7 7 4 4 5 1 5 7 1 1 5 1 1 5 5 5
## [6069] 1 7 7 7 7 7 7 1 5 1 1 1 5 4 4 5 7 7 4 1 1 4 1 1 5 1 1 1 1 7 1 2 1 2 1 1 7
## [6106] 7 4 5 7 4 5 7 5 5 7 5 2 1 7 7 5 4 7 7 5 1 7 5 7 1 1 5 5 7 1 7 4 5 7 1 5 1
## [6143] 1 5 1 1 1 7 7 5 5 5 1 7 7 7 7 1 5 1 7 7 7 4 7 7 7 7 1 1 7 1 5 1 1 2 5 1 1
## [6180] 1 5 1 1 5 5 5 1 5 5 7 5 1 5 4 5 7 7 5 1 1 2 7 5 5 7 5 7 7 7 2 5 1 1 1 1 1
## [6217] 5 1 5 1 5 5 7 5 7 4 7 7 7 7 1 7 1 5 2 5 7 7 1 1 7 5 4 1 5 1 1 1 1 1 1 4 5
## [6291] 5 1 7 1 7 2 5 1 7 1 4 7 5 7 1 5 7 7 7 5 7 1 1 1 1 1 1 1 5 5 7 7 1 1 5 1 1
## [6328] 1 5 1 1 1 1 5 1 5 1 5 1 5 1 2 7 1 1 1 1 5 7 1 1 4 7 4 1 1 5 5 1 1 5 1 5 5 7
## [6365] 5 1 1 1 5 4 7 1 1 1 7 5 7 1 1 5 5 4 7 4 7 5 4 1 1 5 1 1 1 5 1 1 7 7 5 5 5
## [6402] 1 1 1 7 7 5 5 5 7 7 7 1 5 4 7 7 7 1 1 1 1 1 5 5 1 2 4 7 7 7 5 4 7 5 5 1 5 7
## [6439] 7 5 5 1 1 5 1 5 4 7 7 1 1 1 7 7 1 5 1 5 5 5 1 1 5 1 5 5 2 5 5 1 5 5 7 4
## [6476] 4 7 7 7 5 5 7 7 7 5 1 7 4 5 5 7 7 5 7 7 7 7 1 5 5 5 5 5 4 1 7 7 7 5 7 1 5
## [6513] 1 5 7 7 5 5 1 7 2 1 5 7 7 4 4 2 1 7 4 5 1 5 2 2 7 7 1 5 5 1 7 1 4 7 7 5 7
## [6587] 5 1 1 1 7 1 5 7 5 4 7 4 7 1 1 5 5 7 1 5 7 5 7 5 7 1 1 5 1 7 5 1 5 5 2 1 5
## [6624] 5 7 7 2 1 5 4 7 2 1 1 4 7 5 1 5 5 5 1 5 7 5 7 4 7 4 7 5 4 7 7 7 7 7 5 7 5
## [6698] 1 1 7 7 7 4 7 1 7 5 1 7 7 5 7 4 4 7 5 5 1 1 7 5 1 1 5 1 1 7 5 5 4 7 7 7 7
## [6772] 4 7 7 5 1 7 5 1 1 1 5 7 7 7 5 7 1 5 7 5 1 1 7 7 4 4 4 4 7 4 7 7 7 1 1 7 7
## [6809] 4 4 4 7 4 5 4 1 5 2 4 4 7 4 5 5 1 1 6 1 5 1 1 1 7 7 1 5 1 1 2 5 4 1 1 5 1
## [6846] 7 1 7 7 1 2 5 1 1 1 1 1 1 1 5 5 5 7 7 7 5 1 7 2 1 5 7 1 4 7 1 2 4 7 2 5 2
```

```
## [6883] 1 5 2 7 7 1 7 5 5 1 1 1 1 1 5 4 5 4 5 2 7 4 5 1 1 7 1 4 1 1 1 7 7 4 7 5 1
## [6920] 4 5 1 5 1 5 5 1 1 1 5 1 7 7 7 7 7 4 7 1 7 5 1 1 5 1 5 1 7 7 5 7 7 7 7 5 1
## [6957] 5 1 1 1 1 1 5 1 5 1 1 5 1 5 1 5 7 7 7 4 7 7 7 1 5 5 1 1 5 1 1 1 1 5 5 4 1 5 1
## [6994] 5 5 1 4 7 1 1 1 7 5 7 7 1 4 4 7 7 5 1 1 7 1 1 1 1 5 7 7 7 5 5 5 5 7 1 5 1
## [7068] 4 5 1 7 7 7 5 4 7 5 4 5 7 7 7 4 1 5 4 7 1 1 7 1 1 1 5 5 5 7 5 5 5 7 7 4 1
## [7105] 5 7 5 1 1 1 1 1 5 7 2 7 1 7 1 7 4 5 1 1 1 1 1 5 1 7 5 1 2 1 1 1 4 7 4 7 1
## [7142] 5 5 7 4 7 7 1 1 7 1 4 7 4 5 5 1 1 1 7 7 1 5 5 7 1 1 7 5 7 1 1 5 2 7 4 7 4
## [7179] 4 1 1 7 5 5 7 7 7 4 4 7 7 4 5 5 1 1 5 1 1 1 7 5 5 4 1 1 5 7 7 7 1 5 7 1 5
## [7216] 1 1 7 1 7 4 1 7 7 5 1 5 5 5 1 1 1 1 5 1 7 1 4 5 5 5 5 1 2 5 5 7 4 5 1 1 7
## [7253] 1 1 5 1 1 5 5 5 1 5 7 7 5 7 4 7 7 5 7 1 7 4 5 5 5 7 1 5 2 4 4 5 7 1 1 1 7
## [7290] 5 7 5 4 5 1 7 5 5 4 4 5 4 5 7 7 7 1 7 1 1 5 4 4 5 7 1 7 7 4 5 5 7 7 1 1 5
## [7327] 7 1 2 2 4 7 4 7 7 1 7 1 1 1 1 5 1 5 4 7 7 7 4 5 5 5 5 5 5 5 7 7 5 7 1 7 7
## [7401] 1 5 1 5 1 7 7 5 5 7 1 7 7 7 7 5 7 5 4 5 4 7 4 7 7 7 4 1 1 1 1 1 7 4 7 5 5
## [7438] 5 5 5 7 1 7 7 7 7 4 5 5 7 7 4 5 7 7 7 1 1 5 5 7 1 5 1 5 1 5 1 1 1 4 7 2 7
## [7475] 2 1 1 1 1 1 1 1 7 5 1 5 5 2 7 1 5 5 7 7 4 7 1 1 1 1 1 7 1 7 1 4 7 1 7 5 5
## [7512] 5 5 5 1 5 1 5 1 1 2 4 4 7 5 5 5 1 5 7 7 7 7 7 5 1 4 1 7 4 4 7 5 5 5 5 5 5 4
## [7586] 1 1 1 1 5 5 7 7 7 5 1 7 7 7 5 4 5 5 1 4 4 5 7 1 5 1 1 1 1 1 4 1 5 7 5 5 7
## [7623] 1 1 5 5 5 7 5 5 5 5 5 1 1 1 1 1 5 1 5 4 4 4 5 7 4 7 7 5 7 7 7 1 1 1 1 1 1 4
## [7660] 5 4 4 5 5 1 1 2 1 5 1 1 1 1 5 7 1 7 1 1 2 1 5 1 2 5 1 2 1 5 1 1 1 5 7 5 5
## [7697] 4 1 1 1 5 7 1 1 1 1 1 1 7 4 5 7 7 7 5 7 5 4 1 1 1 1 7 7 7 7 7 7 7 7 7 7 7
## [7771] 1 4 3 7 1 1 1 1 7 1 1 4 7 5 7 1 4 7 4 5 4 4 7 5 1 7 5 1 1 1 7 1 7 7 7 7 4
## [7845] 7 5 2 4 7 7 7 7 4 1 1 1 7 5 1 1 4 1 1 7 7 7 4 1 1 1 7 7 1 7 1 7 1 5 7 7 4 7 5
## [7882] 5 5 7 5 7 5 7 5 5 1 5 1 7 5 1 5 5 4 5 7 7 5 7 7 7 4 1 7 5 1 7 4 7 2 1 7 7
## [7919] 7 1 1 1 5 4 1 5 1 1 1 5 1 7 5 1 7 1 7 2 1 5 7 5 4 5 7 4 4 5 5 4 7 7 1 1 1
## [7956] 7 5 1 5 7 7 5 4 7 7 5 1 4 7 7 7 7 7 7 4 7 4 7 1 1 7 7 1 5 5 7 5 5 1 1 1 1 4
## [7993] 7 4 4 4 5 5 5 1 1 5 5 1 1 1 5 1 1 2 5 5 5 5 1 1 5 5 7 5 1 4 5 7 4 1 1 1 1
## [8030] 5 4 5 7 4 4 7 5 1 7 5 5 1 1 1 5 5 4 7 5 4 7 5 1 4 5 7 5 5 7 1 5 7 1 1 1 1
## [8067] 7 1 5 5 4 7 7 4 4 4 1 1 1 5 5 7 5 4 5 5 7 7 7 5 5 1 5 1 1 1 1 5 5 1 1 2 1 2
## [8104] 5 5 7 5 7 7 5 7 7 4 7 5 1 1 5 7 1 7 5 7 7 7 4 7 7 4 1 1 1 1 4 1 5 4 5 5 5
## [8141] 7 1 1 7 7 5 1 1 5 1 1 1 1 6 1 1 7 5 7 5 1 5 1 5 5 7 7 5 2 7 1 5 1 5 1 7 5
## [8178] 5 1 4 5 7 4 4 7 7 7 5 1 1 5 1 4 7 7 1 7 5 7 4 7 5 1 1 2 1 7 7 4 7 7 4 7 4
## [8215] 7 4 4 7 4 7 4 7 7 5 1 5 1 1 1 1 1 5 5 1 7 4 2 7 7 7 7 7 1 5 1 1 4 7 7 4 4
## [8252] 4 7 1 1 5 1 5 5 5 1 1 1 4 1 5 1 7 7 2 7 2 7 7 5 5 4 4 7 1 7 7 7 7 5 4 5 5
## [8289] 5 5 7 7 4 7 1 5 1 7 7 1 1 2 7 7 7 5 5 1 5 7 7 1 2 7 7 5 5 7 7 5 1 7 4 1 5 5
## [8326] 7 4 4 1 1 1 1 5 1 1 5 1 1 5 5 5 1 5 5 7 7 5 1 5 5 1 7 4 7 4 7 1 5 7 5 7 4
## [8363] 5 7 7 7 1 1 5 7 7 4 1 1 5 5 7 7 7 5 7 7 7 1 1 7 1 5 1 5 1 1 1 5 1 7 5 4 7
## [8437] 5 7 7 7 7 7 7 4 7 7 1 1 7 5 1 5 7 7 4 7 7 4 4 4 5 1 1 7 7 1 1 7 2 7 7 7 7
## [8474] 7 4 7 7 1 1 1 1 7 1 5 5 1 1 1 6 5 5 1 5 7 4 7 1 1 1 5 1 7 7 1 7 4 5 7 5 1
## [8511] 1 7 4 7 1 5 1 4 1 1 1 5 1 5 7 5 4 7 7 7 1 1 5 1 1 1 4 5 7 7 5 7 5 1 1 1 1
## [8548] 1 1 7 7 7 7 5 5 5 5 7 5 7 7 4 2 4 5 7 7 5 5 5 7 7 7 1 1 7 1 1 1 1 1 4 7 7 5
## [8622] 7 7 4 5 7 1 5 7 7 1 4 4 7 4 5 5 5 5 1 4 4 5 7 4 7 7 7 5 4 5 1 1 1 1 1 1 1
## [8659] 4 5 5 4 5 5 5 5 5 7 7 1 7 7 7 7 7 7 5 4 7 4 5 2 5 1 1 5 5 1 5 1 7 7 7 7 5 5 1
## [8696] 7 7 7 7 2 5 7 5 7 5 1 1 1 7 7 1 7 7 5 5 4 4 4 5 7 7 5 4 5 5 1 7 7 7 7 5 1
## [8733] 1 1 5 1 1 4 7 7 5 1 7 7 1 1 1 7 7 2 1 7 7 5 4 4 1 5 1 1 2 5 1 1 7 4 4 4 7
## [8807] 1 4 7 5 1 1 5 7 1 5 1 1 5 7 7 4 7 7 7 7 5 5 1 1 5 1 7 1 1 1 1 5 1 7 1 1 1
## [8844] 5 4 7 5 7 7 5 4 7 1 5 7 7 5 7 2 7 1 1 1 1 5 1 7 7 7 5 5 1 4 4 4 7 4 1 1 5
```

```
## [8881] 1 5 7 7 4 4 4 1 5 1 2 1 2 7 1 1 7 1 1 1 4 7 7 7 1 1 5 5 5 1 7 5 4 7 5 4 4
## [8918] 7 4 1 1 7 7 4 7 4 4 5 7 7 7 5 1 1 7 5 5 7 5 7 7 4 4 7 1 1 1 5 1 5 7 7 5 5
## [8955] 7 1 7 4 5 4 5 7 1 1 5 4 7 5 7 5 7 1 7 5 5 7 7 4 1 4 4 4 4 4 4 7 4 7 7 1 2
## [8992] 5 7 4 5 5 7 7 5 7 4 1 1 5 5 5 1 1 1 7 1 7 5 5 5 7 7 1 7 1 5 4 7 1 1 1 1 1
## [9029] 1 7 1 1 1 2 4 7 1 1 7 6 7 7 1 1 7 5 5 5 1 1 1 1 7 7 1 1 2 5 7 5 7 1 5 5 4
## [9066] 7 1 1 7 1 7 1 1 5 5 4 7 1 1 1 5 1 1 1 5 7 4 7 7 5 5 7 4 1 5 1 5 5 5 7 7 7
## [9103] 7 4 4 7 7 7 5 7 5 1 1 1 5 7 7 7 7 5 5 5 1 4 4 4 1 7 4 4 7 1 1 5 7 7 7 7 4 5
## [9140] 5 1 1 7 5 5 4 1 5 2 1 5 7 7 5 7 5 1 5 5 4 4 5 1 7 4 2 1 1 5 1 1 1 1 1 7 1
## [9214] 1 5 5 4 1 5 4 1 2 1 4 4 4 7 5 1 7 5 4 5 1 7 5 7 7 7 5 4 7 5 7 1 5 1 1 7 5
## [9251] 5 7 5 1 1 1 1 1 1 1 1 7 1 5 4 5 1 1 1 5 2 1 1 2 5 5 7 7 4 7 1 5 7 7 1 7 4
## [9288] 7 5 1 1 1 4 5 7 4 7 7 7 4 7 7 7 7 4 5 1 1 1 5 5 7 7 5 1 7 7 7 1 1 1 4 4 5
## [9399] 2 5 1 7 7 2 5 7 7 5 1 4 5 7 2 4 7 1 1 7 7 4 4 5 5 7 1 2 5 5 5 5 7 7 7 5 7
## [9436] 7 4 1 7 7 1 5 7 7 7 1 5 5 1 7 5 1 5 1 7 7 7 4 1 7 1 5 1 7 5 1 5 7 1 1 1 7
## [9473] 1 5 7 5 4 4 5 4 4 7 7 5 5 5 7 1 7 5 2 1 5 2 1 5 1 1 1 5 1 1 5 1 5 1 5 1 1 5 1
## [9510] 5 5 4 7 2 5 5 7 7 7 1 1 7 7 7 5 5 7 5 7 7 1 7 5 5 1 7 4 7 5 1 5 7 4 7 7 1
## [9547] 1 7 4 7 7 5 1 5 5 7 5 4 5 5 7 7 4 7 1 5 4 7 1 4 7 7 7 5 1 7 7 7 1 1 5 5 5
## [9584] 7 1 7 5 7 7 7 4 4 4 1 1 1 7 7 1 7 5 5 5 7 1 7 7 5 1 1 5 1 7 1 5 2 1 7 7 5
## [9621] 5 7 7 4 7 7 7 1 5 4 7 5 5 1 5 2 1 1 1 3 5 5 5 7 5 7 4 1 1 2 5 1 5 7 4 1 7
## [9658] 1 1 5 2 1 7 1 1 1 1 5 1 1 5 5 7 4 7 1 5 1 1 1 5 1 5 5 1 1 5 4 5 1 1 7 5 5
## [9695] 1 7 5 7 1 1 7 1 1 5 1 7 7 1 1 1 7 4 5 5 7 7 7 5 5 7 1 1 1 5 7 1 5 5 1 7 5
## [9732] 5 5 1 1 7 1 1 5 5 1 2 5 1 7 7 7 4 7 1 7 7 7 1 1 1 1 5 1 5 1 1 1 1 5 1 1 5
## [9769] 5 5 7 7 5 1 3 4 4 5 4 4 1 7 5 1 1 1 5 5 7 7 7 5 7 7 4 4 7 7 7 7 7 7 5 7 1 1
## [9806] 7 4 7 4 5 7 1 1 5 5 5 7 5 7 7 5 4 1 1 7 7 7 1 7 7 7 4 7 5 4 4 5 1 1 7 5 1
## [9843] 7 7 4 7 5 5 5 1 1 7 1 5 1 1 5 2 5 1 1 1 1 1 1 1 5 1 7 4 4 7 7 1 7 1 7 4 7
## [9917] 1 1 1 7 4 4 7 7 1 5 1 1 1 2 7 7 5 1 7 7 1 7 5 1 5 5 5 5 1 7 7 2 2 1 1 1 7
## [9991] 1 7 1 1
##
## Within cluster sum of squares by cluster:
## [1] 1131.9138 1960.3599 1460.2766 1749.2333 3515.6513 778.7336 1329.8103
   (between_SS / total_SS = 70.2 %)
##
## Available components:
##
## [1] "cluster"
                   "centers"
                                "totss"
                                                           "tot.withinss"
                                             "withinss"
## [6] "betweenss"
                   "size"
                                "iter"
                                             "ifault"
Re-running k-means with k = 8:
set.seed(123)
clustering_results_8 <- kmeans(scale(data1[,(15:18)]), centers = 8, nstart = 25)
clustering_results_8
## K-means clustering with 8 clusters of sizes 912, 2114, 2923, 12, 9, 182, 3263, 579
## Cluster means:
##
         Sales
                 Quantity
                          Discount
                                         Profit
## 1 -0.21598994 0.03592803 2.6411836 -0.463442556
## 2 0.06144353 0.84777662 -0.3020867
                                    0.051473515
```

```
## 3 -0.12740261 -0.55718427 0.2964566 -0.102300730
## 4 8.36049519 1.14320638 2.0689103 -11.913664966
## 5 16.87915324 0.84359568 -0.6489669 19.798650239
      3.85740734 0.99833967 -0.4172776
                                                       2.749990730
## 7 -0.18963127 -0.56474906 -0.7485896
                                                       0.008899425
## 8 0.17956725 2.49300690 -0.2367609
                                                       0.083089163
##
## Clustering vector:
##
         \begin{smallmatrix} 1 \end{smallmatrix} \begin{smallmatrix} 7 \end{smallmatrix} \begin{smallmatrix} 7 \end{smallmatrix} \begin{smallmatrix} 7 \end{smallmatrix} \begin{smallmatrix} 1 \end{smallmatrix} \begin{smallmatrix} 3 \end{smallmatrix} \begin{smallmatrix} 2 \end{smallmatrix} \begin{smallmatrix} 7 \end{smallmatrix} \begin{smallmatrix} 2 \end{smallmatrix} \begin{smallmatrix} 3 \end{smallmatrix} \begin{smallmatrix} 2 \end{smallmatrix} \begin{smallmatrix} 8 \end{smallmatrix} \begin{smallmatrix} 2 \end{smallmatrix} \begin{smallmatrix} 3 \end{smallmatrix} \begin{smallmatrix} 3 \end{smallmatrix} \begin{smallmatrix} 1 \end{smallmatrix} \begin{smallmatrix} 1 \end{smallmatrix} \begin{smallmatrix} 1 \end{smallmatrix} \begin{smallmatrix} 2 \end{smallmatrix} \begin{smallmatrix} 7 \end{smallmatrix} \begin{smallmatrix} 7 \end{smallmatrix} \begin{smallmatrix} 3 \end{smallmatrix} \begin{smallmatrix} 3 \end{smallmatrix} \begin{smallmatrix} 2 \end{smallmatrix} \begin{smallmatrix} 2 \end{smallmatrix} \begin{smallmatrix} 3 \end{smallmatrix} \begin{smallmatrix} 7 \end{smallmatrix} \begin{smallmatrix} 3 \end{smallmatrix} \begin{smallmatrix} 4 \end{smallmatrix} \begin{smallmatrix} 1 \end{smallmatrix} \begin{smallmatrix} 3 \end{smallmatrix} \begin{smallmatrix} 2 \end{smallmatrix} \begin{smallmatrix} 1 \end{smallmatrix} \begin{smallmatrix} 3 \end{smallmatrix} \begin{smallmatrix} 2 \end{smallmatrix} \begin{smallmatrix} 1 \end{smallmatrix} \end{smallmatrix} \begin{smallmatrix} 3 \end{smallmatrix} \begin{smallmatrix} 2 \end{smallmatrix} \end{smallmatrix} 
##
       ##
##
     [149] \ 7 \ 6 \ 2 \ 2 \ 3 \ 2 \ 7 \ 2 \ 7 \ 3 \ 7 \ 2 \ 7 \ 3 \ 2 \ 3 \ 3 \ 4 \ 3 \ 8 \ 2 \ 1 \ 3 \ 7 \ 7 \ 2 \ 1 \ 3 \ 1 \ 2 \ 2 \ 7 \ 3 \ 2 \ 7 \ 2 \ 7
     [186] 7 8 3 3 2 2 8 3 2 2 3 3 7 1 3 3 1 3 1 2 3 7 7 7 8 7 3 3 1 3 1 3 2 3 3 2 1
##
      \begin{smallmatrix} 223 \end{smallmatrix} \begin{smallmatrix} 3 & 1 & 3 & 7 & 2 & 7 & 3 & 8 & 1 & 3 & 1 & 1 & 3 & 3 & 7 & 2 & 1 & 1 & 3 & 1 & 1 & 3 & 6 & 7 & 2 & 6 & 7 & 3 & 7 & 6 & 2 & 7 & 1 & 3 & 2 & 2 & 7 \\ \end{smallmatrix}
##
##
     [260] 7 2 1 3 8 3 7 3 2 3 3 2 2 7 3 7 7 3 3 2 3 1 2 2 2 2 2 1 3 2 2 8 3 1 8 3 3
      [297] \ 1 \ 3 \ 2 \ 8 \ 7 \ 7 \ 1 \ 7 \ 7 \ 2 \ 7 \ 3 \ 3 \ 3 \ 7 \ 3 \ 3 \ 3 \ 7 \ 6 \ 2 \ 7 \ 7 \ 2 \ 7 \ 3 \ 3 \ 3 \ 8 \ 1 \ 3 \ 1 
##
     [334] \ 1 \ 3 \ 3 \ 3 \ 3 \ 7 \ 3 \ 1 \ 8 \ 8 \ 3 \ 3 \ 7 \ 7 \ 7 \ 7 \ 3 \ 7 \ 6 \ 2 \ 7 \ 7 \ 2 \ 2 \ 8 \ 7 \ 7 \ 7 \ 7 \ 8 \ 7 \ 7 \ 7 \ 2 \ 2
##
     [371] 2 3 1 8 3 8 3 2 1 3 1 2 3 7 3 3 1 3 3 3 7 3 6 1 3 7 7 7 3 3 3 3 3 2 7 2 3
##
##
     [408] \ 2 \ 2 \ 7 \ 8 \ 7 \ 8 \ 3 \ 2 \ 3 \ 8 \ 2 \ 7 \ 7 \ 2 \ 7 \ 7 \ 2 \ 7 \ 8 \ 7 \ 2 \ 1 \ 2 \ 2 \ 3 \ 2 \ 1 \ 2 \ 3 \ 2 \ 8 \ 7 \ 7 \ 7 \ 7 \ 7
##
     [445] \ 7 \ 2 \ 7 \ 7 \ 7 \ 2 \ 7 \ 7 \ 3 \ 3 \ 7 \ 3 \ 3 \ 7 \ 8 \ 3 \ 2 \ 1 \ 1 \ 3 \ 3 \ 1 \ 3 \ 1 \ 7 \ 7 \ 3 \ 3 \ 2 \ 2 \ 2 \ 2 \ 7 \ 2 \ 2
     [482] \ 7 \ 7 \ 7 \ 7 \ 2 \ 2 \ 6 \ 3 \ 3 \ 2 \ 7 \ 7 \ 3 \ 7 \ 8 \ 7 \ 3 \ 8 \ 1 \ 1 \ 3 \ 1 \ 3 \ 2 \ 7 \ 7 \ 7 \ 5 \ 7 \ 7 \ 3 \ 7 \ 7 \ 6 \ 7 \ 7
##
     ##
     [556] 3 7 3 2 7 2 2 3 7 7 3 8 2 7 3 2 7 2 7 7 8 7 7 3 1 7 3 3 3 1 3 7 7 3 3 1 3
##
##
     [593] 3 3 7 3 2 8 3 3 3 3 3 3 1 1 3 1 8 3 3 3 3 8 8 1 7 7 8 7 7 1 7 2 7 8 2 7 2
      [630] \ 7\ 3\ 2\ 3\ 3\ 3\ 7\ 1\ 3\ 2\ 3\ 7\ 7\ 2\ 7\ 7\ 2\ 3\ 7\ 3\ 7\ 8\ 7\ 8\ 3\ 1\ 3\ 3\ 1\ 2\ 8\ 2\ 1\ 2\ 7\ 3
     ##
     ##
     ##
##
     ##
     ##
     [852] \ 2 \ 2 \ 7 \ 7 \ 7 \ 8 \ 7 \ 7 \ 7 \ 7 \ 2 \ 3 \ 3 \ 7 \ 7 \ 3 \ 1 \ 1 \ 3 \ 7 \ 7 \ 2 \ 2 \ 7 \ 7 \ 7 \ 1 \ 3 \ 2 \ 7 \ 7 \ 2 \ 7 \ 3 \ 8 \ 2
##
      [889] \ 2 \ 2 \ 3 \ 7 \ 7 \ 8 \ 2 \ 3 \ 3 \ 2 \ 8 \ 1 \ 2 \ 1 \ 3 \ 7 \ 7 \ 7 \ 3 \ 2 \ 3 \ 2 \ 6 \ 7 \ 2 \ 7 \ 8 \ 3 \ 3 \ 1 \ 3 \ 3 \ 7 \ 7 \ 7 \ 2 \ 3 
     ##
      [963] \ 7\ 3\ 7\ 7\ 7\ 7\ 7\ 2\ 3\ 1\ 3\ 2\ 3\ 1\ 6\ 1\ 7\ 2\ 3\ 3\ 2\ 3\ 3\ 3\ 1\ 2\ 7\ 2\ 3\ 7\ 2\ 6\ 7\ 7\ 7\ 2 
## [1000] 2 7 6 2 7 7 3 3 1 2 3 7 7 7 8 7 2 7 7 2 2 7 7 6 3 3 3 3 3 2 7 7 8 2 3 1 1
## [1037] 7 3 2 3 2 7 3 2 1 8 2 3 8 7 3 3 8 7 7 3 3 3 3 3 3 3 3 2 3 3 3 1 3 3 1 2 3 3
## [1074] 1 1 3 2 7 7 3 7 8 8 7 7 6 7 3 7 3 3 8 3 2 7 2 1 2 7 7 7 7 1 1 7 2 1 1 8 3
## [1111] 2 3 1 3 3 7 7 2 2 7 7 2 2 7 7 7 7 7 7 2 7 7 2 7 3 3 3 8 7 7 2 7 7 7 7 7 3 7 7
## [1148] 7 7 7 7 2 7 2 7 6 7 7 3 7 7 7 7 3 2 7 8 7 7 7 7 7 2 1 2 7 7 7 7 3 2 7 3 2
## [1185] 2 2 7 7 7 8 3 7 7 3 3 3 3 8 3 1 3 1 3 3 3 2 7 3 2 2 3 3 3 1 1 3 3 3 3 2 8
## [1259] 3 8 3 1 3 7 2 7 7 2 3 7 3 1 1 3 2 3 3 3 7 7 7 3 1 3 1 2 7 7 3 7 3 2 7 2 7
## [1296] 8 3 3 2 7 2 7 7 7 7 2 2 3 7 7 3 7 7 3 2 8 2 3 3 1 1 3 2 3 7 1 2 7 7 7 3 3
## [1333] 1 1 8 7 7 7 7 8 1 1 1 7 2 7 2 2 8 2 3 7 2 7 3 3 8 7 1 3 2 3 3 8 1 2 3 3 2
## [1370] 1 3 7 7 3 2 7 8 2 2 2 3 1 7 8 2 7 7 7 7 7 7 7 2 3 1 1 3 7 7 7 7 7 3 2 3 3
## [1407] 3 7 7 1 3 7 7 2 7 7 2 3 1 3 1 3 1 2 2 3 3 2 2 8 8 2 7 8 2 2 3 3 2 2 7 3 7
## [1444] 3 8 3 2 3 1 3 1 2 7 2 6 7 7 8 3 7 3 3 3 2 7 3 3 3 7 7 7 3 7 1 2 3 2 2 3 7
## [1481] 2 7 3 3 7 7 2 7 7 3 7 7 2 7 7 2 3 2 2 3 8 3 3 7 7 2 3 7 7 7 8 8 3 1 1 3 2
## [1518] 3 3 2 2 2 2 8 2 2 3 3 2 7 3 3 3 2 3 7 7 7 7 7 2 2 2 3 7 2 3 1 3 3 1 7 2 2
## [1555] 7 7 3 2 7 8 2 7 1 3 2 3 3 1 7 7 2 7 7 3 3 2 3 7 7 7 8 7 3 7 7 2 7 7 7 2
## [1629] 2 1 3 3 3 3 3 3 3 2 3 7 7 7 7 3 7 6 7 3 3 1 3 7 7 2 7 7 3 7 2 3 3 3 3 3 1 2 3
## [1666] 3 3 3 3 3 3 7 3 7 1 1 7 7 3 3 3 1 7 7 3 1 1 3 3 2 2 2 2 3 7 7 7 7 2 2 3 6
```

```
## [1703] 7 2 1 1 2 3 7 3 2 8 2 1 3 3 2 2 7 8 3 3 3 3 3 3 3 2 7 7 8 3 3 3 3 2 3 7 7
## [1740] 2 8 3 3 3 2 7 1 3 3 3 3 7 8 7 7 8 2 1 3 3 3 2 3 7 7 8 7 1 1 3 1 3 2 7 8 7
## [1777] 7 8 2 7 7 7 7 7 2 8 7 7 7 7 2 7 6 7 3 7 8 7 2 2 3 3 7 1 4 3 6 2 2 2 1 3 2 2
## [1814] 3 3 7 7 7 1 1 1 3 3 3 1 3 3 3 2 3 8 7 3 3 2 2 2 3 3 3 3 7 3 1 7 2 7 3 3 7
## [1851] 2 2 2 7 2 7 2 1 7 7 3 7 3 1 3 3 2 2 3 7 2 3 3 7 3 3 3 3 3 1 7 2 7 2 3 7 3
## [1888] 1 3 3 7 7 3 3 7 3 6 3 1 3 2 7 3 3 3 8 1 1 7 7 7 2 3 2 3 3 3 2 7 2 7 3 8 3
## [1925] 7 7 3 3 7 7 3 3 7 2 2 8 7 1 2 1 3 3 3 1 3 3 7 7 3 7 8 7 2 2 7 7 7 7 7 7 2
## [1962] 2 7 2 7 7 7 2 7 7 3 2 1 3 3 7 2 2 3 2 3 7 7 2 7 7 3 7 2 7 7 1 2 3 3 1 1 7
## [1999] 7 2 2 2 7 2 1 2 2 3 3 3 3 3 3 7 7 2 3 7 7 2 6 7 7 2 2 7 3 3 7 2 7 7 3 3 7
## [2036] 7 7 7 3 1 1 2 2 2 3 7 7 7 2 1 1 8 2 2 7 2 6 7 7 7 2 3 8 3 7 3 8 7 7 3 1 2
## [2073] 3 7 2 7 2 2 2 2 3 3 2 3 7 2 2 7 7 7 7 3 2 7 2 8 7 7 3 2 3 3 3 7 7 3 1 7 3
## [2110] 7 2 3 7 2 7 3 3 2 7 2 1 3 7 8 2 2 3 3 3 7 8 2 7 1 2 3 1 1 7 8 7 3 2 3 2 3
## [2147] 1 2 7 2 3 2 7 3 1 2 7 7 3 1 3 7 7 3 3 1 3 1 3 3 3 3 3 3 1 3 7 8 2 8 2 7 2 6
## [2184] 2 8 3 7 2 7 8 7 3 7 7 7 2 3 7 3 3 2 2 3 7 3 3 1 3 7 2 8 2 7 3 8 7 3 3 3 7
## [2221] 2 3 8 3 3 1 3 3 3 3 7 2 2 3 2 7 8 2 1 3 2 3 7 7 7 7 7 3 7 2 2 2 7 7 7 7 7
## [2258] 2 7 7 7 7 7 7 7 7 7 3 7 7 3 2 2 2 2 3 3 3 2 7 7 3 8 7 2 7 2 3 7 2 2 7 7 2 7 7
## [2332] 2 2 7 2 7 2 7 7 3 2 2 7 3 8 2 1 2 2 3 7 7 2 2 2 7 3 2 1 3 2 7 1 7 7 7 3 7
## [2369] 2 3 7 1 3 8 3 7 1 3 2 3 3 2 7 8 3 3 3 7 3 7 2 8 7 7 7 7 2 7 2 2 2 2 2 2 2 8
## [2406] 2 8 2 3 8 3 2 7 3 2 7 2 3 6 7 7 3 3 2 1 3 8 1 1 1 8 7 2 3 7 7 8 1 3 3 2 3
## [2443] 2 2 2 7 7 2 7 2 2 2 3 7 8 8 7 2 7 7 7 3 2 3 7 7 7 2 2 7 8 3 3 3 2 2 2 3 7
## [2480] 3 3 2 3 2 2 7 7 7 7 2 7 2 6 2 2 8 7 7 3 3 3 1 1 7 7 6 8 1 3 2 2 2 1 3 3 3
## [2517] 8 7 2 2 7 7 7 7 8 3 1 1 3 3 2 7 2 7 2 7 3 3 3 2 3 2 3 7 7 7 3 3 6 8 3 2 7
## [2554] 7 2 7 1 3 2 2 7 3 3 7 7 7 2 6 8 3 3 3 2 2 7 2 7 7 7 8 7 7 8 7 7 3 2 7 7 3
## [2591] 1 1 7 8 7 7 7 7 7 7 7 2 7 7 3 7 1 7 3 3 2 3 3 1 3 3 2 7 8 7 3 7 2 7 5 6 2 7
## [2628] 3 1 1 7 7 3 3 3 1 7 7 3 7 2 7 2 3 8 7 7 7 2 7 2 3 3 7 7 3 7 2 7 2 6 7 7 3
## [2665] 3 1 3 3 3 7 7 2 7 2 7 2 3 1 1 2 2 7 7 7 8 3 1 1 2 3 1 3 3 2 3 3 1 4 3 3 3
## [2702] 3 3 7 2 7 7 7 2 6 1 3 8 3 2 7 7 7 7 1 3 2 2 2 3 3 3 1 2 7 2 7 7 7 7 7 2 7
## [2739] 7 7 7 2 7 2 3 3 3 3 1 3 8 3 7 7 2 6 7 7 7 7 2 7 2 1 1 3 3 6 1 2 3 2 3 7 3
## [2776] 7 2 3 8 3 7 8 7 3 3 3 7 1 2 1 7 7 7 8 3 2 3 7 7 2 2 3 7 7 2 3 7 7 2 3 8 2
## [2813] 7 2 7 2 7 7 2 2 3 3 3 3 8 3 1 3 2 2 3 2 3 2 7 7 7 2 1 7 2 2 3 3 8 7 1 2 6
## [2850] 1 3 7 7 7 7 2 2 3 7 7 6 2 3 7 2 7 7 2 3 3 3 3 3 7 8 7 3 2 7 7 7 2 7 1 7 3 2
## [2924] 7 7 7 7 7 1 7 3 3 3 3 3 7 2 2 2 3 7 6 7 8 7 2 7 8 2 3 2 3 3 3 3 7 7 1 3 7
## [2961] 7 8 3 3 2 7 1 3 3 2 1 3 3 2 3 3 2 7 2 3 3 3 3 3 1 7 7 8 7 3 2 2 7 2 3 2 3
## [2998] 3 3 1 3 1 2 3 7 7 2 3 7 2 1 4 1 1 3 3 8 1 2 7 7 7 2 3 8 7 3 3 2 3 7 7 3 3
## [3035] 2 2 7 2 7 3 2 7 7 6 1 2 7 2 7 7 7 7 7 7 7 6 7 8 7 3 3 8 7 7 8 8 2 2 2 3 6
## [3072] 7 2 7 3 3 1 3 8 7 3 7 7 3 1 3 1 1 2 2 2 2 2 7 3 3 8 1 3 3 3 3 2 7 1 2 3 2 7
## [3109] 7 7 8 7 3 2 2 2 2 7 2 3 3 3 1 1 3 3 3 7 7 3 2 3 7 3 1 3 3 3 3 1 7 1 2 2 2
## [3146] 2 3 2 2 7 6 4 3 2 7 3 3 7 2 3 1 7 2 7 8 3 7 2 8 8 2 1 2 3 2 1 3 7 7 7 2 7
## [3183] 7 6 7 1 3 2 7 3 3 7 1 1 2 7 7 1 3 2 3 3 8 7 7 8 2 8 2 8 3 1 8 3 3 2 7 3 7
## [3220] 2 7 7 3 3 3 2 3 3 3 7 2 1 1 3 3 3 7 3 2 2 1 2 8 2 3 7 2 7 3 3 2 3 3 3 3 3
## [3257] 3 3 1 1 3 7 2 7 3 3 3 7 7 7 7 8 2 6 7 2 2 7 7 2 6 3 3 7 1 7 7 3 2 2 7 7 7
## [3294] 7 1 6 1 3 6 1 3 2 3 1 3 3 3 1 3 2 2 2 2 7 7 3 1 8 7 2 3 2 7 7 1 1 3 2 8 3
## [3331] 3 3 3 1 7 3 7 3 8 7 2 7 3 7 8 7 2 2 1 3 7 3 3 7 3 7 7 8 1 3 7 7 7 7 1 3 3
## [3368] 3 7 3 2 2 7 7 2 7 7 8 1 3 8 3 7 2 6 3 3 3 1 3 3 2 3 1 2 7 7 7 7 2 7 7 1 1
## [3405] 7 3 3 1 3 3 8 3 7 7 3 7 7 2 3 7 7 8 2 3 3 7 8 3 3 2 7 7 8 3 2 1 1 1 3 2 3
## [3442] 8 7 6 3 2 7 7 3 2 7 7 3 7 3 3 3 3 1 1 3 3 2 3 3 7 7 8 2 3 7 3 3 7 7 2 2 8
## [3479] 7 2 2 7 2 7 7 7 3 3 3 2 2 3 3 7 2 2 3 2 2 7 7 8 3 3 2 3 1 3 3 3 3 2 8 3 3
## [3516] 7 8 3 3 1 3 2 7 8 2 7 7 8 3 3 2 2 3 7 2 7 7 7 7 2 2 3 3 3 8 1 2 7 7 8 3 2
## [3553] 8 1 7 2 2 3 3 7 3 1 1 3 7 7 6 1 7 6 6 3 3 7 8 3 3 7 7 2 6 7 7 7 7 7 7 3 7
## [3590] 7 8 3 7 1 2 3 3 2 3 3 8 3 2 2 7 7 7 3 3 3 3 3 3 3 3 2 3 1 3 3 2 3 8 2 7 8 7
## [3627] 7 7 6 3 2 7 8 3 2 3 3 3 2 3 2 1 3 7 7 2 7 7 1 2 2 2 7 7 8 3 3 3 3 3 2 7 8
## [3664] 7 2 1 3 2 8 2 3 3 3 1 3 3 7 7 7 7 7 7 7 8 2 2 7 1 2 2 1 3 2 3 2 2 7 7 2 7
```

```
## [3701] 7 3 2 7 7 2 3 1 1 2 3 2 7 7 7 2 2 7 2 2 7 7 8 1 3 3 7 7 7 7 2 7 2 7 7 7 7
## [3738] 7 7 7 7 7 8 3 7 2 7 7 7 3 2 3 7 3 2 2 2 2 2 7 7 7 2 7 7 1 3 3 2 8 3 2 3 7 3
## [3775] 3 3 2 2 3 7 7 3 3 3 3 1 2 3 3 2 2 1 2 7 7 3 7 3 3 2 3 3 3 3 3 3 2 1 1 2 3
## [3812] 3 7 2 2 3 7 7 2 3 3 7 2 2 3 8 8 6 7 7 7 7 7 3 3 3 3 3 3 8 7 3 3 7 7 2 2 7
## [3849] 2 2 2 7 1 7 2 1 2 3 2 1 3 3 3 7 1 7 2 8 3 2 7 7 3 3 1 3 3 1 3 3 3 3 3 2 3 3
## [3886] 1 7 7 7 7 7 2 8 2 1 2 3 3 3 3 3 3 8 2 3 1 3 3 3 3 3 1 3 7 7 1 3 3 1 3 7 7 3
## [3960] 3 2 2 7 7 3 3 7 1 1 8 8 7 7 2 8 2 7 7 3 3 7 7 3 6 7 3 6 2 2 7 3 3 3 2 7 2
## [3997] 2 7 2 2 8 2 3 3 3 7 7 2 3 8 3 3 7 7 7 2 2 7 8 7 3 3 2 2 3 3 2 3 1 7 7 2 2
## [4034] 7 1 2 3 2 7 2 3 3 8 2 2 1 1 7 8 7 7 3 3 8 3 2 3 2 2 3 3 3 3 8 3 2 3 8 8 1
## [4071] 1 3 7 2 3 3 7 7 7 3 8 1 1 3 1 3 3 3 8 7 7 3 3 6 7 7 2 2 5 7 1 1 8 7 2 3 1
## [4108] 7 3 7 7 7 7 3 3 2 3 7 2 8 3 2 7 3 7 3 2 2 1 7 1 3 3 8 7 2 2 2 1 3 2 7 8 2
## [4145] 7 3 7 1 3 3 3 2 3 3 1 1 3 1 3 7 7 7 7 7 2 8 7 7 3 2 3 2 7 7 7 7 7 2 7 2 8 1
## [4182] 7 7 7 7 2 1 1 3 7 5 2 7 3 7 2 7 3 1 7 3 2 2 2 7 7 7 7 6 3 3 7 7 7 7 7 7 8
## [4219] 6 7 2 2 7 7 1 1 7 3 7 1 1 3 2 1 3 8 1 2 7 7 7 3 7 7 7 7 2 2 2 2 2 2 3 3 3
## [4256] 1 1 1 3 7 2 7 7 1 8 1 2 1 2 2 8 7 8 7 3 1 7 5 7 2 7 7 7 7 3 3 3 3 1 3 7 8
## [4293] 7 3 1 7 7 6 3 2 3 2 3 2 7 2 3 3 2 3 3 7 3 7 7 3 7 7 2 3 7 1 3 1 3 3 7 2 8
## [4330] 3 2 7 2 2 3 3 3 7 7 3 7 7 2 2 7 7 7 7 1 8 3 3 3 7 1 1 3 3 1 3 7 7 7 7 7 2 7
## [4367] 8 3 2 2 7 2 7 2 2 2 7 2 7 7 7 3 1 3 3 2 7 3 2 3 7 7 3 2 3 1 3 3 3 2 7 2 3
## [4404] 7 3 3 7 7 7 7 2 7 7 7 7 3 2 8 3 1 2 3 7 7 2 2 7 7 6 7 7 7 2 7 7 1 2 1 3 3
## [4478] 7 7 2 2 7 8 7 3 7 7 1 8 1 3 3 3 2 2 8 2 1 7 7 2 2 2 2 7 7 3 7 7 3 7 7 1 3
## [4515] 7 7 2 2 7 8 7 7 3 2 2 7 2 7 3 1 7 7 3 7 7 7 3 7 7 7 2 1 7 8 3 2 1 3 1 3 7
## [4552] 3 3 7 7 7 7 1 3 3 7 3 2 3 2 3 1 3 1 7 7 7 3 2 7 2 7 3 7 8 2 3 8 7 2 2 2 7
## [4589] 7 2 7 3 3 7 3 7 8 7 2 7 8 7 3 3 3 2 7 7 3 2 8 7 3 2 1 8 3 7 7 6 3 2 3 3 3
## [4626] 3 2 7 2 7 7 2 7 2 7 7 7 7 7 7 6 7 1 3 3 3 7 7 7 8 2 2 7 7 2 7 3 3 3 3 2 3 2
## [4663] 3 3 1 3 7 7 1 2 2 2 2 7 7 3 3 7 8 7 3 8 8 3 1 1 2 8 1 3 3 3 2 3 7 7 7 8 7
## [4700] 3 3 3 1 1 2 7 7 2 7 7 7 1 2 1 3 3 7 7 2 2 7 1 7 7 1 3 2 3 2 3 2 7 7 7 7 7
## [4774] 2 3 7 7 7 2 7 3 1 2 2 2 1 3 3 8 2 1 3 3 7 7 7 1 2 3 2 3 1 1 3 8 7 8 7 3 7
## [4811] 7 3 3 3 7 3 3 3 7 1 2 1 1 3 3 3 7 7 3 2 7 7 7 2 7 3 1 2 3 7 1 2 7 3 3 2
## [4848] 8 3 7 2 8 2 7 7 7 3 7 8 2 6 2 3 3 7 6 8 7 7 3 2 3 1 3 1 2 3 3 1 7 2 6 7 2
## [4885] 7 2 7 2 7 2 7 7 7 7 7 7 7 2 3 3 1 7 7 2 7 3 7 2 1 3 3 2 3 1 1 7 3 2 3 7 7 1
## [4922] 7 7 2 7 7 3 7 7 2 7 7 3 1 3 3 2 2 3 2 8 7 7 7 7 3 7 7 2 7 7 7 7 7 2 7 7 1
## [4959] 8 3 7 3 7 7 1 3 7 7 7 7 3 3 3 3 7 8 2 3 2 3 1 3 7 7 2 2 3 2 1 3 3 4 8 8 2
## [4996] 3 3 1 3 2 7 2 7 7 7 2 6 7 1 2 1 1 1 2 2 2 3 7 7 7 7 7 7 7 7 7 7 7 7 3 7 7 8
## [5033] 7 7 2 7 7 2 3 7 2 7 7 8 2 2 3 2 1 3 7 3 7 2 3 8 2 2 1 7 8 7 7 7 7 3 2 2 7
## [5070] 7 7 2 7 7 3 2 2 3 7 3 7 3 3 3 7 1 2 3 1 3 3 3 7 1 1 3 1 1 3 3 3 3 3 3 2 3 3
## [5107] 7 8 2 2 2 3 1 3 3 3 7 7 2 8 2 7 7 7 8 2 6 7 7 7 8 8 3 2 3 2 3 7 7 2 7 2 8
## [5144] 3 7 7 7 3 3 7 7 2 3 7 7 7 3 3 1 3 3 8 7 3 7 7 7 7 7 7 7 6 7 7 3 1 3 3 3 7 7
## [5181] 7 2 7 7 2 2 8 7 6 2 3 7 2 2 7 7 7 2 6 3 3 2 2 7 7 7 7 7 7 7 7 7 7 7 7 7 2 2 7
## [5292] 2 3 3 8 3 7 8 7 7 6 6 2 3 1 2 7 7 3 2 4 3 3 7 7 8 3 3 3 1 1 1 3 7 2 2 7 7
## [5329] 3 3 7 2 7 7 7 7 3 2 2 2 7 2 3 2 7 2 3 3 7 2 8 3 3 8 2 7 2 7 7 7 8 2 2 7 3
## [5366] 3 3 7 7 3 7 7 3 2 3 2 7 2 3 7 2 2 8 3 3 8 8 3 7 7 3 2 7 2 2 2 2 8 1 2 7 7
## [5403] 3 3 2 7 7 7 3 3 7 2 7 2 1 3 3 7 2 3 3 3 3 2 7 6 2 2 7 3 3 3 1 3 6 7 3 1 3
## [5440] 3 3 3 7 8 7 3 2 2 7 2 7 7 2 2 3 3 2 7 2 8 3 7 1 3 2 1 3 3 3 7 2 2 3 1 7 7
## [5514] 7 7 7 6 8 3 3 2 3 2 7 8 7 7 7 2 3 6 2 7 7 7 7 7 7 7 7 3 1 3 1 1 1 2 7 3 7 2
## [5551] 3 2 3 2 3 7 7 7 7 2 7 7 6 7 3 3 2 3 3 3 3 3 1 3 3 3 2 7 7 2 8 7 2 7 7 7 7
## [5588] 2 3 7 7 7 7 1 7 7 7 7 7 3 2 3 3 2 2 3 3 2 7 7 3 7 8 7 2 2 2 3 7 2 7 7 8 7
## [5625] 7 7 6 2 3 8 7 2 7 2 2 2 7 7 2 3 8 7 7 7 2 2 8 7 2 2 7 2 7 1 3 3 3 1 3 3
## [5662] 3 8 3 1 3 3 7 8 7 7 3 3 3 7 2 7 1 2 7 7 7 7 2 2 3 2 2 2 3 8 2 7 2 7 3 1 1
```

```
## [5699] 3 2 3 2 3 3 3 7 3 7 7 7 6 7 3 7 2 2 7 2 7 2 2 3 3 7 3 8 2 7 7 7 7 7 8 7 7
## [5736] 3 7 7 3 7 7 3 7 7 2 2 3 3 3 2 1 7 2 2 7 1 6 3 1 7 3 7 7 8 2 3 2 2 7 7 7 2
## [5773] 7 3 2 3 3 7 7 3 3 7 7 7 1 2 1 2 2 1 2 3 7 7 3 8 7 7 7 3 3 3 3 3 3 3 3 3 3 3
## [5810] 3 7 7 3 7 3 7 3 3 8 7 7 7 2 3 3 7 7 7 7 7 7 7 7 3 2 1 1 2 3 2 2 7 3 2 7 2 2
## [5847] 3 7 7 7 7 7 7 7 2 3 1 1 3 7 3 7 7 7 7 7 2 3 1 2 2 7 7 2 7 3 2 7 2 7 3 8 8
## [5884] 7 6 3 7 7 7 7 3 3 3 7 7 7 2 7 7 7 8 7 7 7 2 7 7 3 7 7 3 7 7 2 2 7 7 6 2 7
## [5921] 7 7 7 3 2 8 8 7 7 1 3 3 3 3 3 2 3 3 3 8 7 3 7 2 7 7 2 7 7 7 2 3 2 3 1 2
## [5958] 2 7 3 7 8 7 7 7 2 3 7 7 3 7 7 3 7 2 7 7 8 2 7 7 7 7 2 2 7 7 2 7 2 6 2 3 2
## [5995] 3 2 3 3 3 7 7 3 1 1 7 7 7 7 7 7 6 7 3 2 6 7 2 7 7 2 7 7 3 7 8 3 2 2 3 7 3
## [6032] 7 2 7 1 1 7 2 7 7 3 3 3 3 2 3 7 1 1 2 2 2 3 3 1 1 8 7 2 3 7 7 2 7 7 2 8 8
## [6069] 7 3 3 3 3 3 3 7 2 7 7 7 8 1 1 8 3 3 1 7 7 1 7 7 8 7 7 7 7 3 7 6 7 6 7 7 3
## [6106] 3 1 2 3 1 2 3 2 8 3 2 6 7 3 3 8 1 3 3 2 7 3 2 3 7 7 8 2 3 7 3 1 2 3 7 2 7
## [6143] 7 2 7 7 7 3 3 2 2 8 7 3 3 3 3 7 8 2 3 3 3 1 3 3 3 3 7 7 3 7 2 7 7 6 2 7 7
## [6180] 7 2 7 7 8 2 8 7 2 2 3 8 7 2 1 2 3 3 2 7 7 6 3 2 8 3 2 3 3 3 6 2 7 7 7 7 7
## [6217] 8 7 2 7 2 8 3 2 3 1 3 3 3 3 7 3 7 8 6 2 3 3 7 7 3 8 1 7 2 7 7 7 7 7 7 1 8
## [6254] 7 8 3 1 3 7 7 3 7 1 7 2 3 8 7 2 3 8 2 2 7 7 2 2 7 3 3 3 3 3 3 2 2 3 3 3 2 2
## [6291] 8 7 3 7 3 6 2 7 3 7 1 3 2 3 7 2 3 3 3 8 3 7 7 7 7 7 7 7 2 2 3 3 7 7 2 7 7
## [6328] 7 2 2 7 7 7 2 7 2 7 2 7 7 7 6 3 7 7 7 7 2 3 7 7 1 3 1 7 7 2 2 7 7 2 7 2 7 2 2 3
## [6365] 8 7 7 7 2 1 3 7 7 7 3 2 3 7 7 2 8 1 3 1 3 8 1 7 7 2 7 7 7 2 7 7 3 3 2 8 2
## [6402] 7 7 7 3 3 2 2 2 3 3 3 7 2 1 3 3 3 7 7 7 7 2 8 7 6 1 3 3 3 2 1 3 2 2 7 2 3
## [6439] 3 2 2 7 7 2 7 2 1 3 3 7 7 7 3 3 7 2 7 2 7 2 2 7 7 2 7 2 2 6 2 2 7 2 8 3 1
## [6476] 1 3 3 3 2 2 3 3 3 2 7 2 1 8 2 3 3 2 3 3 3 3 7 8 8 2 2 2 1 7 3 3 3 2 3 7 2
## [6513] 7 2 3 3 2 8 7 3 6 7 2 3 3 1 1 6 7 3 1 2 7 2 6 6 3 3 7 2 2 7 3 7 1 3 3 2 3
## [6550] 2 2 7 2 3 3 3 1 3 8 2 7 1 7 3 7 2 1 2 3 3 3 3 7 7 8 8 3 2 2 7 7 7 7 2 7 2
## [6587] 2 7 7 7 3 7 2 3 2 1 3 1 3 7 7 8 2 3 7 2 3 2 3 2 3 7 7 2 7 3 2 7 2 2 6 7 2
## [6624] 8 3 3 6 7 8 1 3 6 7 7 8 3 2 7 2 8 2 7 8 3 2 3 1 3 1 3 2 1 3 2 3 3 3 2 3 8
## [6698] 7 7 3 3 3 1 3 7 3 2 7 2 3 8 3 1 1 3 2 2 7 7 3 2 7 7 2 7 7 3 2 2 1 3 3 3 3
## [6772] 1 3 3 2 7 3 8 7 7 7 8 3 3 3 2 3 7 2 3 2 7 7 3 3 1 1 1 1 3 1 3 3 3 7 7 3 3
## [6809] 1 1 1 3 1 8 1 7 2 6 1 1 3 1 2 8 7 7 5 7 2 7 7 7 3 3 7 2 7 7 6 2 1 7 7 2 2
## [6846] 3 7 3 3 7 6 2 7 7 7 7 7 7 7 2 8 8 3 3 3 2 7 3 6 7 8 3 7 1 3 7 6 1 3 6 2 6
## [6883] 7 2 6 3 3 7 3 8 2 7 7 7 7 7 2 1 2 1 2 6 3 1 2 7 7 3 7 1 7 7 7 3 3 1 3 2 7
## [6920] 1 2 7 2 7 2 8 7 7 7 2 7 3 3 3 3 3 1 3 7 3 2 2 7 2 7 2 7 3 3 2 3 3 3 3 2 7
## [6957] 2 7 7 7 7 7 2 7 2 7 7 8 7 2 3 3 3 1 3 3 3 7 8 2 7 7 2 7 2 7 7 2 2 1 7 2 7
## [6994] 2 2 7 1 3 7 7 7 3 2 3 3 7 1 1 3 3 2 7 7 3 7 7 7 7 2 3 3 3 8 2 8 2 3 7 2 7
## [7031] 7 7 7 3 3 3 3 7 7 7 3 7 2 3 3 3 3 7 3 3 2 7 7 7 2 3 8 2 7 7 3 3 3 3 7 2 2
## [7068] 1 8 7 3 3 3 2 1 3 2 1 8 3 3 3 1 7 8 1 3 7 7 3 7 7 7 2 2 2 3 2 8 2 3 3 1 7
## [7105] 2 3 8 7 7 7 7 7 2 3 6 3 7 3 7 3 1 2 7 7 7 7 7 2 7 3 2 7 6 7 7 7 1 3 1 3 7
## [7142] 2 2 3 1 3 3 7 7 3 7 1 3 1 2 2 7 7 7 3 3 7 2 8 3 7 7 3 8 3 7 7 2 6 3 1 3 1
## [7179] 1 7 7 3 2 8 3 3 3 1 1 3 3 1 8 2 7 7 2 7 7 7 3 1 2 1 7 7 8 3 3 3 7 8 3 7 2
## [7216] 7 7 3 7 3 1 7 3 3 2 7 2 2 2 7 7 7 7 8 7 3 7 1 2 2 2 2 7 6 2 2 3 1 8 7 7 3
## [7253] 7 7 2 7 7 2 2 2 7 2 3 3 2 3 1 3 3 8 3 7 3 1 2 2 8 3 7 2 6 1 1 2 3 7 7 7 3
## [7290] 8 3 8 1 2 7 3 2 2 1 1 2 1 2 3 3 3 7 3 7 7 2 1 1 8 3 7 3 3 1 2 2 3 3 7 7 2
## [7327] 3 7 2 6 1 3 1 3 3 7 3 7 7 7 7 8 7 2 1 3 3 3 1 2 2 2 8 2 2 8 3 3 2 3 7 3 3
## [7364] 3 8 1 3 2 1 3 7 8 7 2 7 3 3 2 3 7 2 2 7 7 7 7 2 8 7 3 2 3 2 7 2 3 7 2 8 3
## [7401] 7 2 7 2 7 3 3 2 8 3 7 3 3 3 3 2 3 2 1 2 1 3 1 3 3 3 1 7 7 7 7 7 7 3 1 3 2 2
## [7438] 2 2 2 3 7 3 3 3 3 1 2 2 3 2 1 2 3 3 3 7 7 2 8 2 7 8 7 8 7 8 7 7 7 1 3 6 3
## [7475] 6 7 7 7 2 7 7 7 3 2 7 2 2 6 3 7 2 2 3 2 1 3 7 7 7 7 7 3 7 3 7 1 3 7 3 8 2
## [7512] 2 2 2 7 2 7 2 7 7 6 1 1 3 2 2 2 7 2 3 3 3 3 3 2 7 1 7 3 1 1 3 2 2 2 2 2 1
## [7549] 1 1 7 2 3 7 1 2 2 7 7 7 2 3 3 3 3 7 3 3 8 7 7 2 2 7 7 7 7 3 7 6 7 3 7 6 2
## [7586] 7 7 7 7 2 8 3 3 3 2 7 3 3 3 2 1 2 8 7 1 1 8 3 7 2 7 7 7 7 7 1 7 2 3 2 2 3
## [7623] 7 7 2 8 2 3 2 2 2 2 8 7 7 7 7 7 2 7 2 1 1 1 2 3 1 3 2 8 3 3 3 7 7 7 7 7 1
## [7660] 2 1 1 2 2 7 7 6 7 2 7 7 7 7 2 3 7 3 7 7 6 7 2 7 6 8 7 8 7 8 7 7 7 2 3 2 2
```

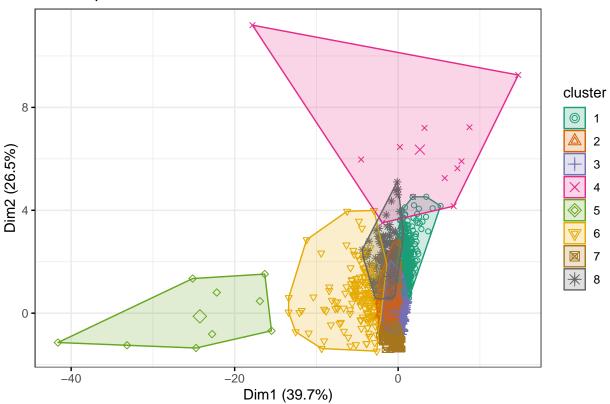
```
## [7697] 1 7 7 7 2 3 7 7 7 7 7 7 3 1 2 3 3 3 2 3 2 1 7 7 7 3 3 3 3 3 3 7 3 3 3 3 3
## [7771] 7 1 4 3 7 7 7 7 3 7 7 1 3 8 3 7 1 3 1 8 1 1 3 2 7 3 2 7 7 7 3 7 3 3 3 3 1
## [7808] 1 3 3 7 7 3 2 3 3 7 2 6 8 2 8 7 7 2 7 3 3 3 3 7 3 3 3 2 3 3 3 7 7 7 2 7 7 8
## [7845] 3 2 6 1 3 3 3 3 1 7 7 7 3 2 7 7 1 7 7 3 3 3 1 7 7 7 3 3 7 2 7 7 2 3 3 1 3 2
## [7882] 2 8 3 2 3 2 3 2 8 7 2 7 3 2 7 8 2 1 2 3 3 2 3 3 3 1 7 3 8 7 3 1 3 6 7 3 3
## [7919] 3 7 7 7 2 1 7 2 7 7 7 2 7 3 2 7 3 7 3 6 7 8 3 8 1 2 3 1 1 2 2 1 3 3 2 7 7
## [7956] 3 2 7 2 3 3 2 1 3 3 8 7 1 3 3 3 3 3 1 3 1 3 7 7 3 3 7 2 2 3 2 2 7 7 7 7 1
## [7993] 3 1 1 1 2 8 2 7 7 2 2 7 7 7 8 7 7 2 2 2 2 8 7 7 2 8 3 8 7 1 2 2 1 7 7 7 2
## [8030] 2 1 2 3 1 1 3 2 7 3 2 2 7 7 7 8 2 1 3 2 1 3 2 7 1 2 3 8 2 3 7 2 3 7 7 7 7
## [8067] 3 7 2 2 1 3 3 1 8 1 7 7 7 8 2 3 2 1 8 2 3 3 3 8 2 7 2 7 7 7 2 2 7 7 6 7 6
## [8104] 2 2 3 2 3 3 2 3 2 1 3 2 7 7 2 3 7 3 8 3 3 3 1 3 3 1 7 7 7 7 1 7 2 1 8 8 8
## [8141] 3 7 7 3 3 2 7 7 2 7 7 7 7 5 7 7 3 2 3 2 7 2 7 2 2 2 3 3 2 6 3 7 2 7 2 7 3 2
## [8178] 2 7 1 8 3 1 1 3 3 3 8 7 7 2 7 1 3 3 7 3 2 3 1 3 2 7 7 6 7 3 3 1 3 3 1 3 1
## [8215] 3 1 1 3 1 3 1 3 3 2 7 8 7 7 7 7 7 2 2 7 3 1 6 3 3 3 3 3 7 8 7 7 1 3 3 1 1
## [8252] 1 3 7 7 8 7 2 2 8 7 7 7 1 7 2 7 3 3 6 3 6 3 3 2 8 1 1 3 7 3 3 3 3 8 1 2 2
## [8289] 2 2 3 3 1 3 7 8 7 3 3 7 7 6 3 3 3 2 2 7 8 3 3 7 6 3 3 2 2 3 3 8 7 3 1 7 2 2
## [8326] 3 1 1 7 7 7 7 2 7 7 8 7 7 2 2 2 7 2 2 3 3 2 7 8 2 7 3 1 3 1 3 7 2 3 2 3 1
## [8363] 2 3 3 3 7 7 2 3 3 1 7 7 2 2 3 3 3 2 3 3 3 7 7 3 7 2 7 2 7 7 7 2 7 3 2 1 3
## [8400] 3 2 7 7 7 7 2 7 2 2 3 3 2 3 8 3 7 2 7 7 7 7 7 2 7 6 1 3 7 2 2 7 8 3 3 1 1
## [8437] 8 3 3 3 3 3 3 1 3 3 7 7 3 8 7 2 3 3 1 3 3 1 1 1 2 7 7 3 3 7 7 3 6 3 3 3 3
## [8474] 3 1 3 3 2 7 7 7 3 7 2 8 7 7 7 5 8 2 7 8 3 1 3 7 7 7 2 7 3 3 7 3 1 2 3 8 7
## [8511] 7 3 1 3 7 8 7 1 7 7 7 8 7 2 3 2 1 3 3 3 7 7 2 7 7 7 1 2 3 3 2 3 2 7 7 7 7
## [8548] 7 7 3 3 3 3 8 2 2 8 3 8 3 3 1 6 1 2 3 3 2 2 8 3 3 3 7 7 3 7 7 7 7 1 3 3 2
## [8622] 3 3 1 2 3 7 2 3 3 7 1 1 3 1 2 2 2 2 7 1 1 8 3 1 3 3 3 2 1 2 7 7 7 7 7 7 7
## [8659] 1 2 2 1 2 2 2 2 2 3 3 7 3 3 3 3 3 2 1 2 1 2 6 2 7 7 2 2 7 2 7 3 3 3 3 2 2 7
## [8696] 3 3 3 3 6 8 3 2 3 8 7 7 7 3 3 7 3 3 2 2 1 1 1 8 3 3 8 1 2 2 7 3 3 3 3 2 7
## [8733] 7 7 2 7 7 1 3 3 2 7 3 3 7 7 7 3 3 6 7 3 3 2 1 1 7 8 7 7 6 8 7 7 3 1 1 1 3
## [8770] 3 3 2 3 2 3 3 2 8 3 3 3 2 3 7 3 2 3 3 3 8 2 7 2 2 3 7 7 2 6 3 7 7 7 7 3
## [8807] 7 1 3 2 7 7 2 3 7 2 7 7 2 3 3 1 3 3 3 3 2 2 7 7 8 7 3 7 7 7 7 2 7 3 7 7 7
## [8844] 2 1 3 2 3 3 8 1 3 7 2 3 3 2 3 6 3 7 7 7 7 2 7 2 3 3 2 8 7 1 1 1 3 1 7 7 8
## [8881] 7 2 3 3 1 1 1 7 8 7 6 7 6 3 7 7 3 7 7 7 1 3 3 3 7 7 8 2 8 7 3 8 1 3 2 1 1
## [8918] 3 1 7 7 3 3 1 2 1 1 2 3 3 3 2 7 7 3 8 2 3 2 3 3 1 1 3 7 7 7 2 7 2 3 3 2 2
## [8955] 3 7 3 1 2 1 2 3 7 7 2 1 3 2 3 8 3 7 3 2 2 3 3 1 7 1 1 1 1 1 1 3 1 3 3 7 6
## [8992] 2 3 1 2 2 3 3 2 2 1 7 7 2 8 2 7 7 7 3 7 3 2 2 2 3 3 7 3 7 2 8 3 7 7 7 7 7
## [9029] 7 3 7 7 7 6 1 3 7 7 3 5 3 3 7 7 3 2 2 2 7 7 7 7 3 3 7 7 6 2 3 2 3 7 2 8 1
## [9066] 3 7 7 3 7 3 7 7 2 2 1 3 7 7 7 2 7 7 7 2 3 1 2 3 2 8 3 1 7 2 7 2 2 2 3 3 3
## [9103] 3 1 1 3 3 3 2 3 2 7 7 7 2 3 3 3 3 2 2 2 7 1 1 1 7 3 1 1 3 7 7 2 3 3 3 1 2
## [9140] 2 7 7 3 8 2 1 7 2 8 7 2 3 3 2 3 2 7 8 8 1 1 2 7 3 1 6 7 7 8 7 7 7 7 7 3 7
## [9177] 1 7 7 3 3 3 7 7 7 3 6 3 3 2 1 3 1 3 2 3 7 2 2 3 7 7 3 2 8 7 8 3 7 7 7 7 7
## [9214] 7 2 2 1 7 2 1 7 6 7 1 1 1 3 2 7 3 8 1 8 7 2 2 3 3 3 2 1 3 2 3 7 2 7 7 3 2
## [9251] 8 3 2 7 7 7 7 7 7 7 7 7 3 7 2 1 2 7 7 7 2 6 7 2 6 2 2 3 3 1 3 7 2 3 3 7 3 1
## [9288] 3 2 7 7 7 1 2 3 1 3 3 3 1 3 3 3 3 1 2 7 7 7 2 8 3 3 2 7 3 3 3 7 7 7 1 1 8
## [9399] 6 2 7 3 3 6 2 3 3 8 7 1 2 3 6 1 3 7 7 3 3 1 1 2 2 3 7 6 2 2 8 2 3 3 3 2 3
## [9436] 3 1 7 3 3 7 2 3 3 3 7 2 2 7 3 8 7 2 7 3 3 3 1 7 3 7 2 7 3 2 7 2 3 7 7 7 3
## [9473] 7 2 3 2 1 1 2 1 1 3 3 2 8 8 3 7 3 2 6 7 8 6 7 2 7 7 7 8 7 7 8 7 2 7 7 2 7
## [9510] 2 8 1 3 6 2 8 3 3 3 7 7 3 3 3 2 2 3 2 3 3 7 3 8 2 7 3 1 3 2 7 2 3 1 3 3 7
## [9547] 7 3 1 3 3 2 7 2 2 3 2 1 2 2 3 3 1 3 7 2 1 3 7 1 3 3 3 2 7 3 3 2 7 7 2 2 2
## [9584] 3 7 3 2 3 3 3 1 1 1 7 7 7 3 3 7 3 8 2 2 3 7 3 3 2 7 7 8 7 3 7 2 6 7 3 3 2
## [9621] 2 3 3 1 3 3 3 7 2 1 3 2 8 7 2 8 7 2 7 4 2 2 2 3 2 3 1 7 7 6 8 7 2 3 1 7 3
## [9658] 7 2 2 6 7 3 7 7 7 7 2 7 7 2 2 3 1 3 7 2 7 7 7 8 7 2 8 7 7 2 1 2 7 7 3 8 2
```

```
## [9695] 7 3 2 3 7 7 3 7 7 2 7 3 3 7 7 7 7 3 1 2 2 3 3 3 2 2 3 7 7 7 8 3 7 2 2 7 3 2
## [9732] 8 8 7 7 3 7 7 2 2 7 6 2 7 3 3 3 1 3 7 3 3 3 3 7 7 7 7 2 7 2 7 7 7 7 2 7 7 2 7 7 2
## [9769] 2 2 3 3 2 7 4 1 1 8 1 1 7 3 2 7 7 7 2 2 3 3 3 2 3 3 1 1 3 3 3 3 3 2 3 7 7
## [9806] 3 1 3 1 2 3 7 7 2 2 2 3 2 3 3 8 1 7 7 3 3 3 3 7 3 3 3 1 3 2 1 1 2 7 7 3 8 7
## [9843] 3 3 1 3 2 2 2 7 7 3 7 2 7 7 2 6 2 7 7 7 7 7 7 7 7 2 7 3 1 1 3 3 7 3 7 3 1 3
## [9880] 2 3 7 8 7 7 7 7 7 7 7 7 2 2 7 2 8 3 3 3 3 3 3 3 3 1 3 2 3 3 2 7 7 2 3 7 7 2
## [9917] 7 7 7 2 1 1 3 3 7 8 7 7 7 6 3 2 2 7 3 3 7 3 2 7 2 8 8 2 7 3 3 6 6 7 7 7 3
## [9954] 7 3 3 2 7 7 3 7 3 3 3 7 7 2 2 7 2 7 2 3 2 2 7 3 2 3 8 7 2 2 7 8 7 7 7 2 3
## [9991] 7 3 7 7
##
## Within cluster sum of squares by cluster:
## [1] 1692.9327 1850.0638 1220.0787 1460.2766 778.7336 1798.3840 994.1753
## [8] 958.1986
   (between_SS / total_SS = 73.1 %)
##
## Available components:
##
## [1] "cluster"
                      "centers"
                                     "totss"
                                                     "withinss"
                                                                    "tot.withinss"
## [6] "betweenss"
                      "size"
                                     "iter"
                                                     "ifault"
```

The within cluster sum of squares by cluster value is 73.1% for k=8, which is not very different from the Within cluster sum of square by cluster value for k=7 (70.2%).

Let's plot the K-means clusters

## Cluster plot



There are still some overlaps between cluster groups.

Compute PCA and extract individual components and extract individual components.

```
# Dimension reduction using PCA
results_pca_8 <- prcomp(data1[,(15:18)], scale = TRUE)

# Coordinates of individuals
ind.coord <- as.data.frame(get_pca_ind(results_pca_8)$coord)

# Add clusters obtained using the K-means algorithm
ind.coord$cluster <- factor(clustering_results_8$cluster)

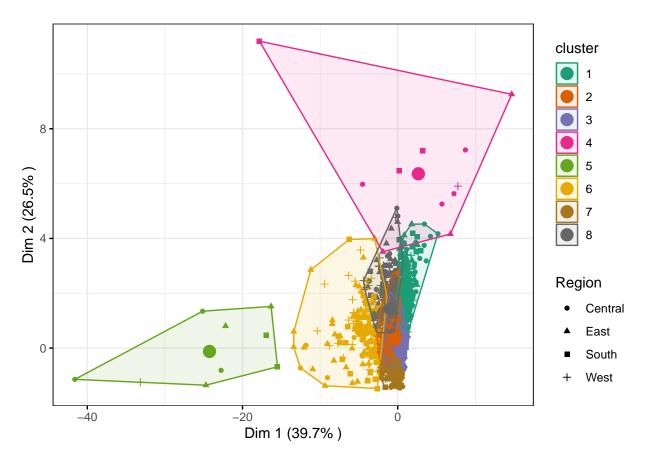
# Add Region groups from the original data sett
ind.coord$Region <- data1$Region
#look at the first few rows to double check
head(ind.coord)</pre>
```

```
##
         Dim.1
                    Dim.2
                                Dim.3
                                            Dim.4 cluster Region
## 1 -0.0426607 -1.0508481 0.14635987 0.31244832
                                                        7
                                                           South
## 2 -1.1546701 -0.7511273 0.25351937
                                                          South
                                       0.22331213
## 3 0.3107253 -1.1081474 -0.01276717
                                       0.15000936
                                                        7
                                                            West
## 4 0.6283963 1.9637213 0.37261495
                                      1.54367992
                                                        1 South
## 5 0.5768386 -0.4056909 0.55980463 -0.06649362
                                                        3
                                                           South
## 6 -0.3936310  0.3152509 -1.56442060 -0.20148637
                                                            West
```

```
# Percentage of variance explained by dimensions
eigenvalue <- round(get_eigenvalue(results_pca_8), 1)
variance.percent <- eigenvalue$variance.percent
head(eigenvalue)</pre>
```

```
eigenvalue variance.percent cumulative.variance.percent
##
## Dim.1
                                 39.7
## Dim.2
                1.1
                                 26.5
                                                              66.2
## Dim.3
                                                              88.3
                0.9
                                 22.0
## Dim.4
                0.5
                                 11.7
                                                             100.0
```

```
ggscatter(
  ind.coord, x = "Dim.1", y = "Dim.2",
  color = "cluster", palette = brewer.pal(n = 8, name = "Dark2"), ellipse = TRUE, ellipse.type = "conversable = "Region", size = 1.5, legend = "right", ggtheme = theme_bw(),  #change point s
  xlab = paste0("Dim 1 (", variance.percent[1], "%)"),
  ylab = paste0("Dim 2 (", variance.percent[2], "%)")
) +
  stat_mean(aes(color = cluster), size = 4) #add cluster centroid using stat_mean()
```



# Clustering Validation

(https://towardsdatascience.com/clustering-analysis-in-r-using-k-means-73eca4fb7967)

Silhouette coefficient can be used to evaluate the goodness of the clustering. First, for each observation i, it calculates the dissimilarity between i and all the other points within the same cluster. This value is called average dissimilarity Di.

Then calculate the dissimilarity between i and all the other clusters and get the lowest value among them. Find the dissimilarity between i and the next closest cluster, called Ci.

Next find the silhouette width which is the difference between Ci and Di, divided by the maximum difference between Ci and Di. Si = (Ci - Di)/max(Di, Ci)

Si>0 means the observation is well clustered. The closer it is to 1 the better it is clustered.

Si < 0 means the observation is wrongly clustered.

Si = 0 means the observation is between 2 clusters.

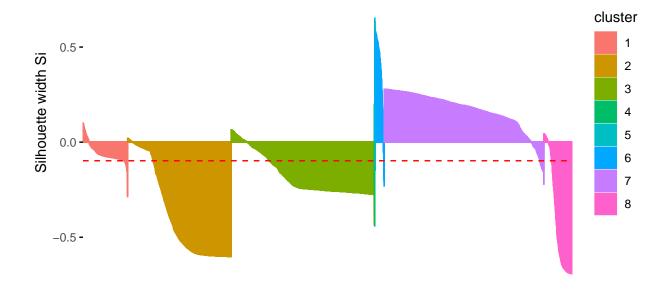
```
library(cluster)
library(factoextra)

sil <- silhouette(clustering_results_8$cluster, dist(data1[,(15:18)]))
fviz_silhouette(sil)</pre>
```

##		cluster	size	ave.sil.width
##	1	1	912	-0.06
##	2	2	2114	-0.38
##	3	3	2923	-0.18
##	4	4	12	-0.17
##	5	5	9	0.48
##	6	6	182	0.37
##	7	7	3263	0.17
##	8	8	579	-0.40

# Clusters silhouette plot Average silhouette width: -0.1

1.0 -



From the table, 5 of the 8 clusters have a negative silhouette width which means that some observations may be in the wrong cluster, so the clustering is not very good.

### Method 2: K-medoids

Following this tutorial: https://towardsdatascience.com/clustering-on-mixed-type-data-8bbd0a2569c3

K-means algorithm is limited in that it can only work with numerical data, whereas our dataset contains both numeric and categorical data.

We will now try the PAM clustering algorithm (Partitioning across medoids). K-medoids is more robust to outliers and noise than K-means. This algorithm uses Gower distance to measure the partial dissimilarity across individuals, and ranges in [0 1]. Standardization is first applied to the features, and the distance between individuals represents the average of all feature specific distances.

Partial dissimilarity is different depending on the type of variable: numeric or categorical.

Numeric features - partial dissimilarity is dependent on absolute difference between 2 observations (x\_i and x\_j), and the maximum range observed from all individuals.  $d_ij^f = |x_i - x_j|/|\max_N(x) - \min_N(x)|$  where N is the number of individuals in a dataset.

Categorical/Qualitative features - feature dissimilarity is equal to 1 if y\_i and y\_j do not have the same values, otherwise it is 0.

One method to determine the number of clusters is by using the Silhouette coefficient.

### **Data Preparation**

For this K-medoids analysis we will omit the Region variable because there are only 4 values and the information is too broad.

We will try to cluster transactions according to the following features:

Compute Gower distance

```
gower_dist <- daisy(data2, metric = "gower")
gower_mat <- as.matrix(gower_dist)

#Print most similar transactions
data2[which(gower_mat == min(gower_mat[gower_mat != min(gower_mat)]), arr.ind = TRUE)[1, ], ]

## Ship_Mode Segment City State Sub_Category diff_in_days Sales
## 4875 Standard Class Consumer Houston Texas Binders 4 1.188
## 3326 Standard Class Consumer Houston Texas Binders 4 1.234
## Quantity Discount Profit</pre>
```

## **Deployment**

## 4875

## 3326

## Responsible ML Framework

1

1

### Conclusion

### References

```
https://www.datanovia.com/en/blog/k-means-clustering-visualization-in-r-step-by-step-guide/
```

https://www.tidymodels.org/learn/statistics/k-means/

0.8 - 1.9602

0.8 - 1.9744

https://towardsdatascience.com/clustering-analysis-in-r-using-k-means-73eca4fb7967

https://towardsdatascience.com/clustering-evaluation-strategies-98a4006fcfc

https://www.analyticsvidhya.com/blog/2020/12/a-case-study-to-detect-anomalies-in-time-series-using-anomalize-package-in-r/