

2. 分担研究報告書

木村 昌臣

データマイニングを使った バーコードアンケートデータの解析

木村 昌臣 (芝浦工業大学)

May 17, 2017

1 はじめに

医薬品の安全管理に関するバーコードの利用状況について収集されたアンケートについて、データマイニング手法を適用して解析を行った。

アンケートの解析の際には、特に項目間（同一設問の項目間、複数設問にわたる項目間など）の関係導出に着目した。

アンケートは、医療機関（病院等）、保険薬局、薬局を対象に行われたが、本稿では、回答が多く収集された医療機関（病院等）、保険薬局を対象に解析を行った結果についてまとめた。

2 方法

2.1 医療機関

2.1.1 医療機関についてのクラスタリング

回答した医療機関についてグルーピングを行うため、「I 医療機関に関する項目」の回答に対してクラスタリングを適用した。なお、クラスタリングに利用した各項目の意味は以下の通りである。

- q1:病院種別(1:一般病院、2:療養型病院、3:精神科病院、4:それ以外(ケアミックス))
- q2.1:特定機能病院(1:あり、2:なし)
- q2.2:地域医療支援病院(1:あり、2:なし)
- q2.3:へきち医療拠点病院(1:あり、2:なし)
- q2.4:臨床研修指定病院(1:あり、2:なし)
- q2.5:がん診療連携拠点病院(1:あり、2:なし)

- q2.6:災害拠点病院(1:あり、2:なし)
- q3:病院の規模(病床数)(1:20~49床、2:50~99床、3:100~299床、4:300~499床、5:500床以上)

未回答は-1とし、q1~q3を未回答ありの回答を除外した上で階層的クラスタリングを適用した。クラスター数は得られたデンドログラムより目視にて決定し、さらに、クラスターの特徴を探るため、各クラスターにおける変数の平均値・分散について調べた。

2.1.2 バーコードリーダー有無について

さらに、設問「薬剤部門にある医薬品に関するバーコードリーダーの有無について、該当するもの全てにチェックをつけてください。」についてクラスタリング手法を適用し、各クラスターで各部門について有無に関する平均・分散を求め、各クラスターの傾向を求めた。さらに、このクラスターと医療機関についてのクラスターのクロス表を求め、その傾向について調べた。

2.1.3 バーコードシステムの導入・拡大についての決定木分析

目的変数を「バーコードシステムを導入あるいは拡大する計画はありますか」(1:ある、2:検討中である、3:計画はない)の回答とし(ただし、解析では検討中も計画ありに含めることとし、“1”と“2”を“1”にまとめた)、説明変数を「薬剤部門にある医薬品に関するバーコードリーダー有無」のクラスター、医薬品に関するバーコードの利用有無およびq1からq3とした。

2.1.4 バーコードを変更してほしい事例について

「以下のような変更があった場合に共通商品コード（バーコード）を変更した方がよいと思われる場合に○を付けて下さい」という質問に対する回答（複数選択可）の傾向をみるため、再びクラスタリング手法を適用し、得られたクラスタの特徴を各選択肢の平均値・分散を求めることにより得た。さらに、q1～q3と、医薬品のバーコード利用の有無、バーコードシステムを導入あるいは拡大する計画の有無、バーコードリーダーを持つ部門クラスタ、バーコードのチェックを製品ごとに行っているか否かを説明変数に、バーコードを変更してほしい事例のクラスタを目的変数として再び決定木手法を適用した。

2.2 保険薬局

2.2.1 回答した保険薬局のクラスタリング

回答した薬局についてグルーピングを行うため、「I 薬局に関する項目」の回答に対してクラスタリングを適用した。なお、クラスタリングに利用した各項目の意味は以下の通りである。

- q1:処方箋を応需する主たる医療機関の種別(1:病院、2:診療所)
- q2:処方せん枚数(枚/月)
- q3:開設主体(1:個人、2:法人)
- q4:経営薬局数(1:1薬局、2:～10薬局、3:～20薬局、4:～50薬局、5:～99薬局、6:100～)
- q5:薬剤師数(常勤換算数)(1:1人、2:2～3人、3:4～6人、4:7～9人、5:10人～)
- q6:後発医薬品調剤割合(1:0～49%、2:50～64%、3:65～74%、4:75%～)

未回答は-1とし、q1～q6を未回答ありの回答を除外した上で階層的クラスタリングを適用した。

2.2.2 バーコードリーダー有無について

さらに、設問「業務で利用している医薬品に関するバーコードリーダーの有無について、該当するもの全てにチェックをつけてください。」についてクラスタリング手法を適用し、各クラスタで各業務につ

いて有無に関する平均・分散を求め、各クラスタの傾向を求めた。さらに、このクラスタと保険薬局のクラスタのクロス表を求め、その傾向について調べた。

2.2.3 バーコードを変更してほしい事例について

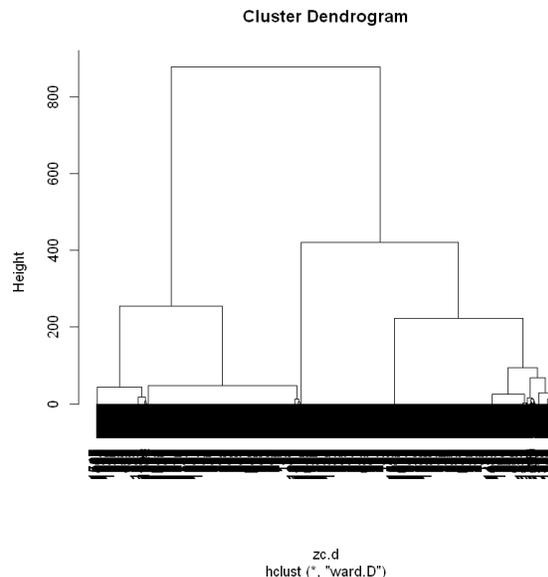
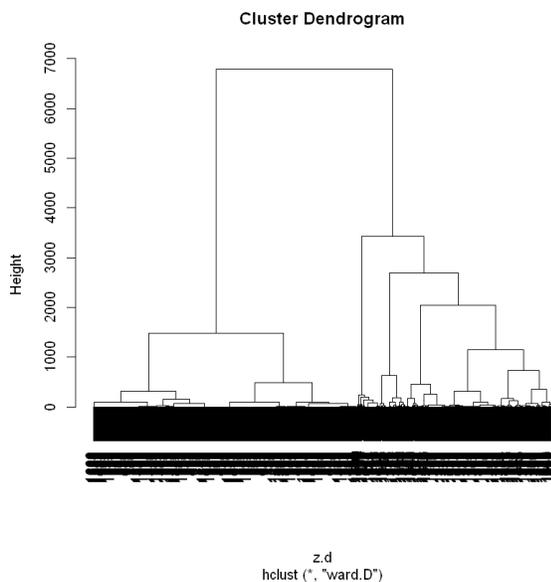
「以下のような変更があった場合に共通商品コード（バーコード）を変更した方がよいと思われる場合に○を付けて下さい」という質問に対する回答（複数選択可）の傾向をみるため、再びクラスタリング手法を適用し、得られたクラスタの特徴を各選択肢の平均値・分散を求めることにより得た。さらに、q1～q6と、医薬品のバーコード利用の有無、バーコードシステムを導入あるいは拡大する計画の有無、バーコードリーダーを持つ業務クラスタを説明変数に、バーコードを変更してほしい事例のクラスタを目的変数として再び決定木手法を適用した。

3 結果

3.1 医療機関

3.1.1 医療機関についてのクラスタリング

クラスタリング手法を適用したところ、次の dendrogram が得られた。



この結果より、データは7つのクラスタに分けられることがわかる。

- クラスタ1:20-49床の承認・指定のないケアミックス等
- クラスタ2:100-299床のへき地医療拠点、臨床研修指定、災害拠点の一般病院等
- クラスタ3:100-299床の一般病院等
- クラスタ4:300-499床の地域医療支援、臨床研修指定、災害拠点の一般病院等
- クラスタ5:300-499床の地域医療支援、へき地医療拠点、がん診療連携拠点、災害拠点の一般病院等
- クラスタ6:500床以上の特定機能、臨床研修指定、がん診療連携拠点、災害拠点の一般病院等
- クラスタ7:100-299床の臨床研修指定の一般病院等

3.1.2 バーコードリーダー有無について

また、バーコードリーダーの有無についてクラスタリングを行うと、以下のような結果となった。

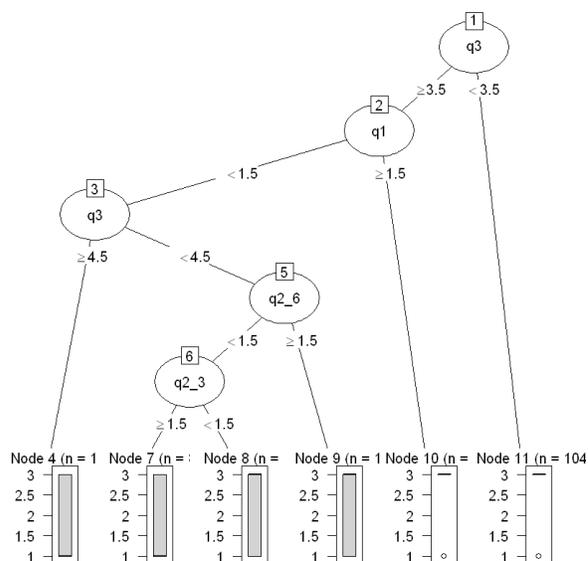
- クラスタ1:薬剤管理部門と調剤部門
- クラスタ2:薬剤管理部門と調剤部門と注射薬調剤部門
- クラスタ3:回答のチェックなし
- クラスタ4:薬剤管理部門と調剤部門と注射薬調剤部門と無菌調剤部門
- クラスタ5:薬剤管理部門のみ
- クラスタ6:調剤部門のみ
- クラスタ7:薬剤管理部門と調剤部門と注射薬調剤部門と製剤部門
- クラスタ8:調剤部門と注射薬調剤部門
- クラスタ9:その他

これらのクラスタと医療機関のクラスタの間でクロス表を求めた。上の表は件数を集計したものであり、下の表は、この件数をもとにリフトを求めたものである。リフトは、クロス表の行・列の関連性の強さを示す量であり、 i を医療機関のクラスタ番号、 j をバーコードリーダー有無のクラスタ番号と

し、セル(i, j)にある頻度を n_{ij} とすると、リフトは次の式で得られる。

$$lift_{ij} = \frac{n_{ij}n_{..}}{n_{i.}n_{.j}}$$

で表される。ただし、 $n_{i.} = \sum_j n_{ij}$ 、 $n_{.j} = \sum_i n_{ij}$ 、 $n_{..} = \sum_{i,j} n_{ij}$ である。この量が1より大きければ共起傾向が強いとみることができる。この結果から、医療機関クラスタ2,5,6とバーコードリーダー有無クラスタ2,4,7,8の間で大きなリフト値が見られる。このことから、指定を受けている医療機関では複数の部門でバーコードリーダーを持っている様子がかがえる。一方で、医療機関クラスタ1,3ではバーコードリーダー有無クラスタ3のリフト値が大きい。これはケアミックスや一般病院でバーコードリーダーが導入されている部門の回答がなかった傾向にあることがうかがえる。



- 300床未満の病院では導入の計画はないという回答が多い傾向があり、
- 300床以上の大きな病院においては
 - － 一般病院でない場合には導入の計画はないという回答が多い傾向があり、
 - － 一般病院の場合、
 - * 500床以上であれば計画はあり、
 - * 500床未満では
 - ・ 災害拠点病院であれば計画はあり
 - ・ 災害拠点病院でなければ計画はないという回答が多い傾向が見られた

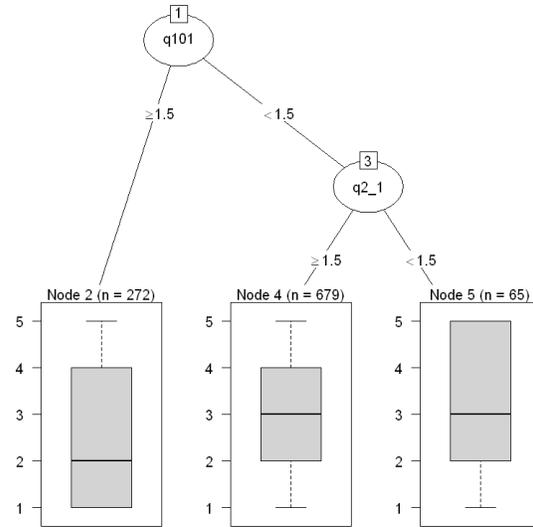
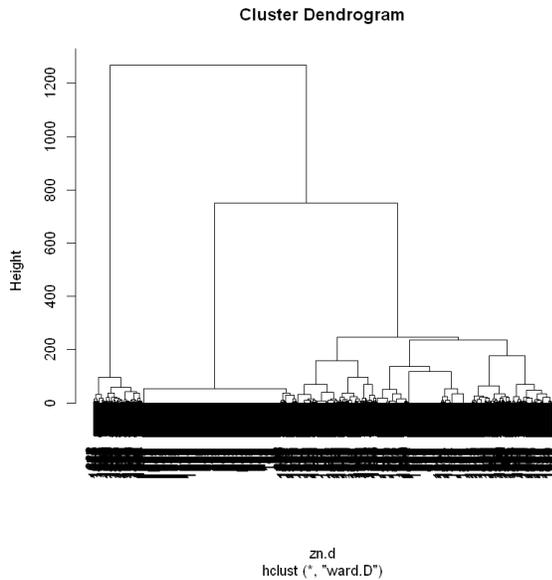
		バーコードリーダー有無クラスタ								
		1	2	3	4	5	6	7	8	9
医療機関 クラスタ	1	81	14	185	1	107	50	4	2	8
	2	23	11	13	5	15	7	3	4	3
	3	61	19	205	2	100	33	1	5	5
	4	49	28	32	14	34	12	5	4	3
	5	47	23	10	19	11	13	12	8	0
	6	14	10	5	13	9	6	11	6	1
	7	47	13	40	5	39	18	3	2	1
		バーコードリーダー有無クラスタ								
		1	2	3	4	5	6	7	8	9
医療機関 クラスタ	1	0.9	0.4	1.3	0.1	1.2	1.2	0.3	0.2	1.3
	2	1.3	1.7	0.5	1.5	0.9	0.9	1.4	2.4	2.6
	3	0.7	0.6	1.5	0.1	1.1	0.8	0.1	0.6	0.8
	4	1.3	2.0	0.6	2.0	0.9	0.7	1.1	1.1	1.2
	5	1.6	2.1	0.2	3.5	0.4	1.0	3.3	2.8	0.0
	6	0.9	1.7	0.2	4.5	0.6	0.9	5.8	4.0	1.0
	7	1.3	1.0	0.7	0.8	1.1	1.2	0.7	0.6	0.4

3.1.3 バーコードシステムの導入・拡大についての決定木分析

バーコードシステムを導入あるいは拡大する計画の有無についての決定木は以下のように得られた。

このことから、病院の規模が導入・拡張有無に大きく影響があることが見て取れる。

3.1.4 バーコードを変更してほしい事例について



このデンドログラムからバーコードを変更してほしい事例についての5つのクラスタがあると考えられ、各クラスタについては以下のように解釈できる。

- クラスタ1: 回答なし
- クラスタ2: 主に剤形・有効成分に関する情報を付した販売名への変更、ブランド名の変更、販売会社変更時、銘柄別⇔統一名称変更時
- クラスタ3: 主に剤形・有効成分に関する情報を付した販売名への変更、ブランド名の変更、製剤の色・形・大きさ・識別記号の変更、販売会社変更時、管理保管方法変更時、銘柄別⇔統一名称変更時
- クラスタ4: 主にブランド名の変更のみ
- クラスタ5: 主に有効成分以外の成分・分量変更時、管理保管方法変更時

これを目的変数として決定木分析を行うと次のような木が得られた。

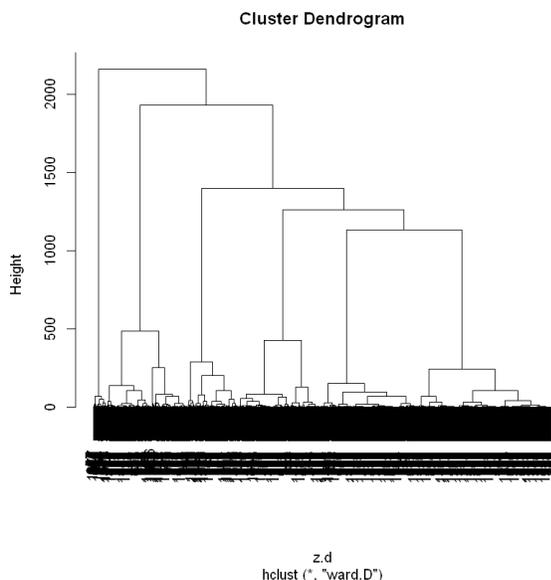
この結果より、以下のように回答されている傾向にあることがわかる。

- バーコードを利用していない場合、クラスタ1 (回答なし)
- バーコードを利用している場合
 - 特定機能病院でない場合、クラスタ4 (主にブランド名の変更のみ)
 - 特定機能病院である場合、クラスタ2 (主に剤形・有効成分に関する情報を付した販売名への変更、ブランド名の変更、販売会社変更時、銘柄別⇔統一名称変更時)

3.2 保険薬局

3.2.1 回答した保険薬局のクラスタリング

クラスタリングの結果、以下のデンドログラムを得た。

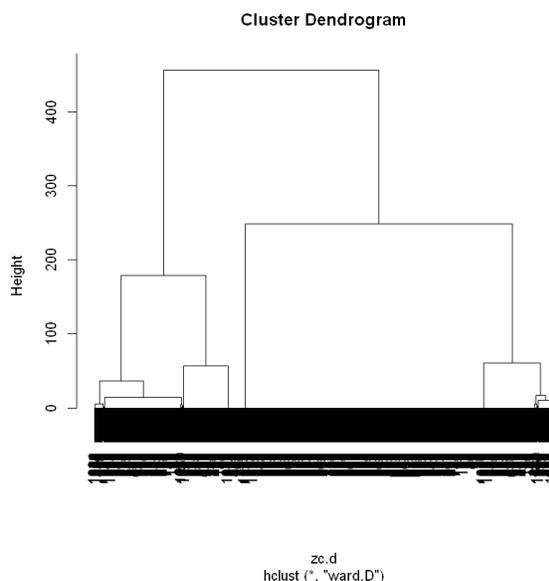


このデンドログラムから、データは6つのクラスターに分けられうるとみなすことができる。各クラスターにおける変数の平均値・分散について調べ、その結果よりそれぞれのクラスターはその特徴を鑑みて以下のように解釈することができる。

- クラスター1:【中小規模】処方箋枚数約1400枚、診療所応需、法人、薬局数21~50、薬剤師2~3人、後発品65~74%
- クラスター2:【病院】処方箋枚数約1200枚、病院応需、法人、薬局数100~、薬剤師2~3人、後発品75%~
- クラスター3:【後発品多】処方箋枚数約1200枚、診療所応需、法人、薬局数100~、薬剤師2~3人、後発品75%~
- クラスター4:【大規模】処方箋枚数約3300枚、診療所応需、法人、薬局数100~、薬剤師4~6人、後発品65~74%
- クラスター5:【後発品少】処方箋枚数約1400枚、診療所応需、法人、薬局数100~、薬剤師2~3人、後発品50~64%
- クラスター6:【個人】処方箋枚数約1600枚、診療所応需、個人、薬局数51~99、薬剤師2~3人、後発品65~74%

3.2.2 バーコードリーダー有無について

業務ごとのバーコードリーダー有無について得られたデンドログラムは次のとおりである。



各クラスターにおける変数の平均値・分散より、それぞれのクラスターの特徴を鑑みて以下のように解釈することができる。

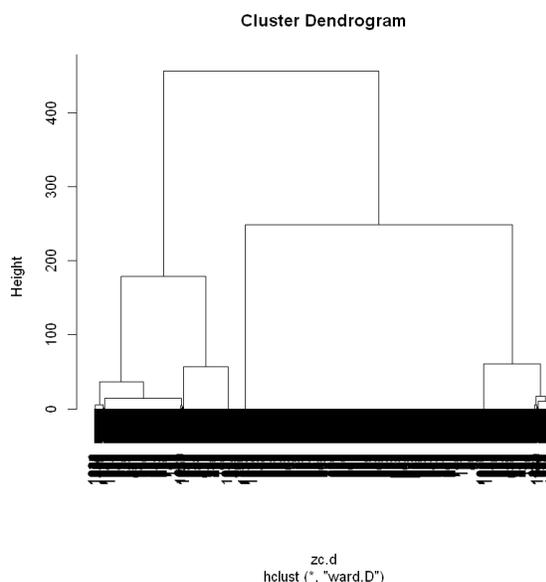
- クラスター1:回答にチェックなし
- クラスター2:薬剤管理業務と調剤業務
- クラスター3:薬剤管理業務と調剤業務と注射薬調剤業務
- クラスター4:薬剤管理業務のみ

これらのクラスターと保険薬局のクラスターとの間で求めたクロス表は下図のとおりである。この結果をみると、リフト値ベースでは保険薬局クラスター1,6とバーコードリーダー有無のクラスター1,4で大きな値が見て取れる。一方でそれ以外の保険薬局クラスターではバーコードリーダー有無のクラスター2,3でリフト値が1を超える傾向がある。このことから、個人もしくは小規模の保険薬局ではバーコードリーダーが使われていない、もしくは薬剤管理業務のみで使われている傾向にあり、規模が大きくなると

薬剤管理業務だけでなく調剤業務など複数の業務でのバーコードリーダーの活用の様子がうかがえる。

		バーコードリーダー有無クラスタ			
		1	2	3	4
保険薬局 クラスタ	1	47	9	7	43
	2	11	127	39	29
	3	26	159	36	52
	4	18	96	31	28
	5	17	104	24	23
	6	9	6	4	8

		バーコードリーダー有無クラスタ			
		1	2	3	4
保険薬局 クラスタ	1	3.3	0.2	0.4	2.1
	2	0.4	1.2	1.3	0.7
	3	0.7	1.1	0.9	1.0
	4	0.8	1.1	1.2	0.8
	5	0.8	1.2	1.0	0.7
	6	2.5	0.4	1.0	1.5



このデンドログラムから3つのクラスタが得られることがわかり、それぞれについては以下のように解釈できる。

クラスタ1:回答にチェックなし (572件)
 クラスタ2:ブランド名の変更、製剤の色・形・大きさ・識別記号の変更、管理保管方法変更時 (188件)

クラスタ3:ブランド名の変更、管理保管方法変更時、銘柄別⇄統一名収載変更時 (193件)
 これにより変更してほしい事例については、バーコード変更に関する全ての回答にチェックがない件数が半数を超えることがわかった。そのため、決定木分析を適用したが、根ノードから分岐しなかった。

3.2.3 バーコードを変更してほしい事例について

バーコードを変更してほしい事例について得られたデンドログラムは次のとおりである。

4 考察・まとめ

医療機関においては、バーコードシステムの導入については病院機能の承認・指定の有無が多くの部門でバーコードリーダーが使われているかどうかに影響していることがわかった。一方で、一般病院やケアミックスでは導入されている部門の回答がなかったケースが特徴的であった。さらに、システムの導入・拡大に関しては医療機関の規模も影響している結果が得られた。バーコード利用の普及に

おいては一般病院やケアミックスにおいてより多くの部門で利用されることが鍵になると考えられる。バーコードの変更に関しては特定機能病院以外はブランド名変更時、特定機能病院ではブランド名に限らず剤型・有効成分・販売会社・銘柄別統一名収載変更時に変更を希望している傾向がある。

保険薬局においては、バーコードシステムの導入については個人・小規模では利用が少なく、規模が大きくなると多くの業務で利用される傾向があった。

A 解析ログ

解析した内容を示すため、実行したRコマンドおよび得られた結果をログとして次ページ以降に示す。

資料 2

資料 2 - 1 医療機関アンケート解析ログ

資料 2 - 2 薬局アンケート解析ログ

医療機関回答データの解析を行う。そのために、まず、医療機関の回答データの読み込みを行う。

In [1]:

```
df <- read.csv("D:/questionair/170418医療機関回答.csv")  
df
```

id	user_id	q1	q2_1	q2_2	q2_3	q2_4	q2_5	q2_6	q3	...	q301_5	q301_6	q301_7
2	1389	3	2	2	2	2	2	2	3	...	0	0	0
3	7293	1	2	2	1	1	2	1	3	...	0	0	0
4	1081	1	2	2	2	2	2	2	4	...	0	0	0
5	1130	1	2	1	2	1	2	1	4	...	0	0	0
6	8438	1	2	1	2	2	2	2	1	...	0	0	0
7	7820	3	2	2	2	2	2	2	4	...	1	1	1
8	3108	1	2	2	2	2	2	2	3	...	0	0	0
9	1726	4	2	2	2	2	2	2	3	...	0	0	0
10	2408	4	2	2	2	2	2	2	3	...	0	0	0
11	1686	1	2	1	2	1	2	2	4	...	0	0	0
12	2155	3	2	2	2	2	2	2	3	...	0	0	0
13	2491	1	2	2	2	1	1	1	5	...	0	0	0
14	1687	3	2	2	2	2	2	2	3	...	0	0	0
15	356	3	2	2	2	2	2	2	3	...	0	0	0
16	1433	2	2	2	2	2	2	2	3	...	1	1	1
17	7847	1	2	2	2	2	2	2	2	...	0	0	0
18	3402	1	2	1	2	1	2	1	4	...	0	0	0
19	2940	1	2	2	2	2	2	2	1	...	0	0	0
20	1724	1	1	1	2	1	1	1	5	...	0	0	0
21	2745	1	2	1	2	1	1	1	4	...	0	0	0
22	3883	4	2	2	2	1	2	2	3	...	0	0	1
23	1327	1	2	2	2	1	2	2	3	...	0	0	0
24	4451	1	2	2	2	2	2	2	3	...	0	0	0
25	2322	1	2	1	1	1	1	1	4	...	0	0	0
26	310	1	2	2	2	1	1	1	4	...	0	0	0
27	8086	4	2	2	2	2	2	2	3	...	0	0	0
28	8273	3	-1	-1	-1	-1	-1	-1	3	...	0	0	0
29	3417	4	2	2	2	2	2	2	2	...	0	0	0
30	348	2	2	2	2	2	2	2	3	...	0	0	0
31	2365	4	2	2	2	2	2	2	3	...	1	0	0
...
1663	4582	-1	-1	-1	-1	-1	-1	-1	-1	...	0	0	0
1664	2518	1	2	2	2	1	2	2	4	...	0	0	0
1665	1784	1	2	1	2	1	2	1	4	...	0	0	0

id	user_id	q1	q2_1	q2_2	q2_3	q2_4	q2_5	q2_6	q3	...	q301_5	q301_6	q301_7
1666	6658	4	2	2	2	2	2	2	2	...	0	0	0
1667	4590	1	2	2	2	2	2	2	3	...	0	0	0
1668	6184	3	2	2	2	2	2	2	3	...	0	0	1
1669	7575	3	2	1	2	2	2	2	3	...	0	0	0
1670	7517	1	2	1	1	1	1	1	4	...	0	0	0
1671	3232	3	2	2	2	2	2	2	4	...	0	0	0
1672	2250	1	2	1	2	1	1	1	5	...	0	0	1
1673	3200	1	2	1	2	1	1	1	5	...	0	0	0
1674	7638	3	2	2	2	2	2	2	3	...	0	0	0
1675	3019	4	2	2	2	2	2	2	3	...	1	1	0
1676	8458	2	2	2	2	2	2	2	2	...	0	0	0
1677	1538	4	2	2	2	2	2	2	4	...	0	0	0
1678	6135	1	2	1	2	1	1	2	3	...	0	0	1
1679	6786	3	2	2	2	2	2	2	3	...	0	1	1
1680	3629	1	2	2	2	1	2	1	4	...	1	0	0
1681	7610	3	2	2	2	2	2	2	3	...	0	0	0
1682	8078	3	2	2	2	2	2	2	3	...	0	0	0
1683	1811	3	2	2	2	2	2	2	3	...	1	1	1
1684	7686	2	2	2	2	2	2	2	2	...	0	0	0
1685	4561	1	2	1	2	1	2	2	4	...	1	1	1
1686	7936	1	2	2	2	2	2	2	2	...	0	0	0
1687	6254	1	2	2	2	2	2	2	1	...	0	0	0
1688	4295	1	2	1	2	1	1	2	4	...	0	0	1
1689	2313	1	2	2	1	2	2	1	3	...	0	0	0
1690	3467	1	2	2	2	1	2	1	5	...	0	0	0
1691	4609	1	1	1	-1	1	1	1	5	...	1	0	0
1692	7090	1	2	1	2	1	2	2	3	...	0	0	0



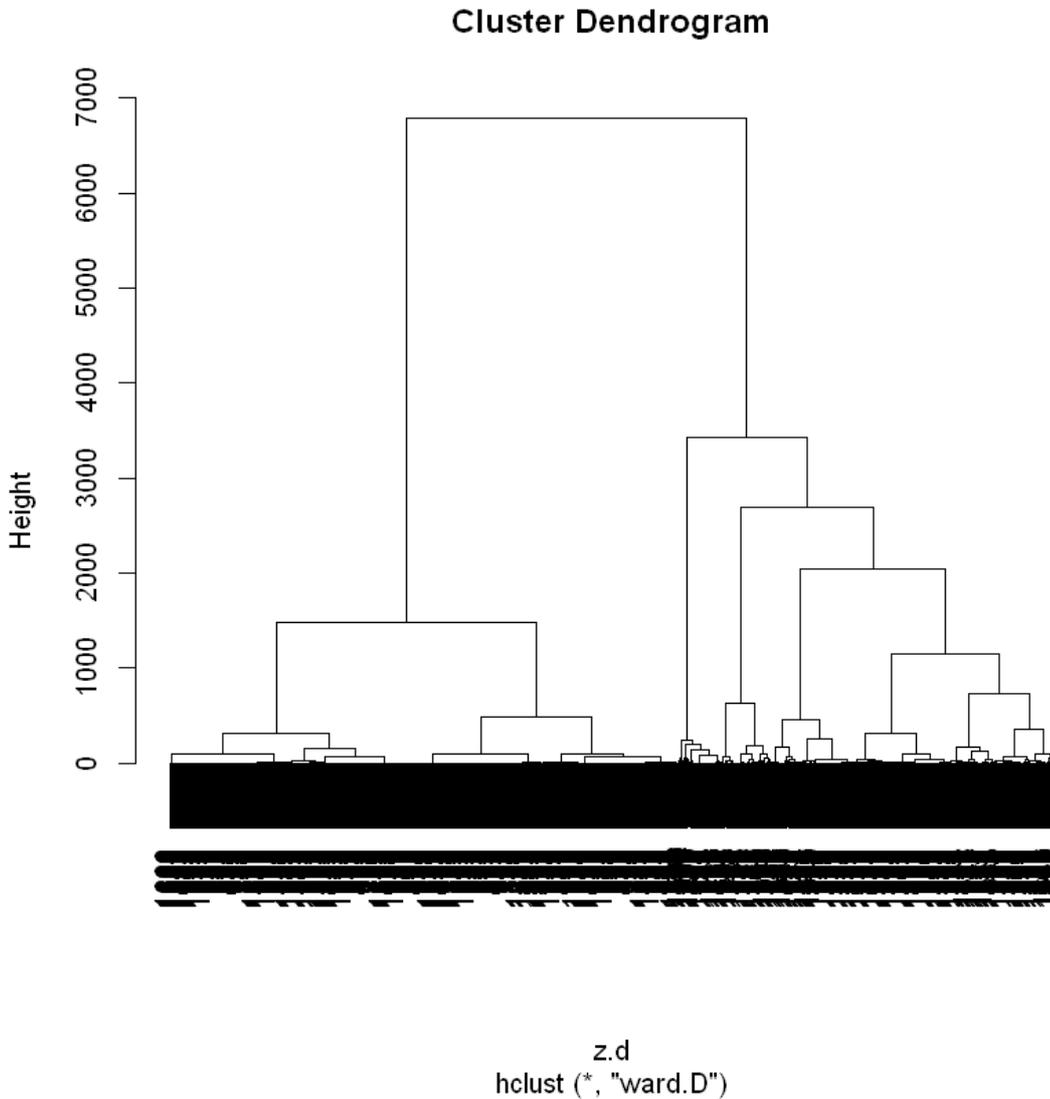
回答した医療機関についてグルーピングを行うため、「I 医療機関に関する項目」の回答に対してクラスタリングを適用する。なお、データフレームの各列の意味は以下の通りである。

- q1:病院種別 (1:一般病院、2:療養型病院、3:精神科病院、4:それ以外 (ケアミックス))
- q2_1:特定機能病院(1:あり、2:なし)
- q2_2:地域医療支援病院(1:あり、2:なし)
- q2_3:へきち医療拠点病院(1:あり、2:なし)
- q2_4:臨床研修指定病院(1:あり、2:なし)
- q2_5:がん診療連携拠点病院(1:あり、2:なし)
- q2_6:災害拠点病院(1:あり、2:なし)
- q3:病院の規模 (病床数) (1:20~49床、2:50~99床、3:100~299床、4:300~499床、5:500床以上)

未回答は-1としている。q1~q3を未回答ありの回答を除外した上で階層的クラスタリングを適用する。

In [2]:

```
df1 <- df[df$q1>0 & df$q2_1>0 & df$q2_2>0 & df$q2_3>0 & df$q2_4>0 & df$q2_5>0 & df$q2_6>0 & df$q3>0, ]
z<-df1[, 3:10]
za<-scale(z)
id<-df1[, 1:2]
z.d <- dist(za)^2
result_cluster <- hclust(z.d, method="ward.D")
plot(result_cluster)
```



得られたグラフ（デンドログラム）の左側に大きなクラスタ二つ、右側に（チェーン気味であるが）段階的に大きく5つに分かれているとみなすことができる。そのため、このデンドログラムからクラスタをk=7個に分割する。

In [3]:

```
c1<-cutree(result_cluster, k=7)
(z1<-cbind(z, c1))
z2<-cbind(id, z1)
z_cluster <- z2[, c(1, 11)]
```

	q1	q2_1	q2_2	q2_3	q2_4	q2_5	q2_6	q3	c1
1	3	2	2	2	2	2	2	3	1
2	1	2	2	1	1	2	1	3	2
3	1	2	2	2	2	2	2	4	3
4	1	2	1	2	1	2	1	4	4
5	1	2	1	2	2	2	2	1	4
6	3	2	2	2	2	2	2	4	1
7	1	2	2	2	2	2	2	3	3
8	4	2	2	2	2	2	2	3	1
9	4	2	2	2	2	2	2	3	1
10	1	2	1	2	1	2	2	4	4
11	3	2	2	2	2	2	2	3	1
12	1	2	2	2	1	1	1	5	5
13	3	2	2	2	2	2	2	3	1
14	3	2	2	2	2	2	2	3	1
15	2	2	2	2	2	2	2	3	1
16	1	2	2	2	2	2	2	2	3
17	1	2	1	2	1	2	1	4	4
18	1	2	2	2	2	2	2	1	3
19	1	1	1	2	1	1	1	5	6
20	1	2	1	2	1	1	1	4	5
21	4	2	2	2	1	2	2	3	7
22	1	2	2	2	1	2	2	3	7
23	1	2	2	2	2	2	2	3	3
24	1	2	1	1	1	1	1	4	2
25	1	2	2	2	1	1	1	4	5
26	4	2	2	2	2	2	2	3	1
28	4	2	2	2	2	2	2	2	1
29	2	2	2	2	2	2	2	3	1
30	4	2	2	2	2	2	2	3	1
31	2	2	2	2	2	2	2	4	3
...
1652	1	2	2	2	2	2	2	3	3
1653	1	2	2	2	2	2	2	3	3
1655	1	2	2	2	1	2	2	4	7

	q1	q2_1	q2_2	q2_3	q2_4	q2_5	q2_6	q3	c1
1656	1	2	1	2	1	2	1	4	4
1657	4	2	2	2	2	2	2	2	1
1658	1	2	2	2	2	2	2	3	3
1659	3	2	2	2	2	2	2	3	1
1660	3	2	1	2	2	2	2	3	4
1661	1	2	1	1	1	1	1	4	2
1662	3	2	2	2	2	2	2	4	1
1663	1	2	1	2	1	1	1	5	5
1664	1	2	1	2	1	1	1	5	5
1665	3	2	2	2	2	2	2	3	1
1666	4	2	2	2	2	2	2	3	1
1667	2	2	2	2	2	2	2	2	3
1668	4	2	2	2	2	2	2	4	1
1669	1	2	1	2	1	1	2	3	5
1670	3	2	2	2	2	2	2	3	1
1671	1	2	2	2	1	2	1	4	4
1672	3	2	2	2	2	2	2	3	1
1673	3	2	2	2	2	2	2	3	1
1674	3	2	2	2	2	2	2	3	1
1675	2	2	2	2	2	2	2	2	3
1676	1	2	1	2	1	2	2	4	4
1677	1	2	2	2	2	2	2	2	3
1678	1	2	2	2	2	2	2	1	3
1679	1	2	1	2	1	1	2	4	5
1680	1	2	2	1	2	2	1	3	2
1681	1	2	2	2	1	2	1	5	4
1683	1	2	1	2	1	2	2	3	4

クラスタの特徴を探るため、各クラスタにおける変数の平均値・分散について調べる。(結果のgroup:xは、クラスタxに相当) 結果より、それぞれのクラスタの特徴を鑑みて以下のように解釈することができる。なお、病院種別については、一般病院の件数が他よりも多いことから、クラスタ1以外については一般病院等とした(療養型、精神科も含まれる)。

- クラスタ1:20-49床の承認・指定のないケアミックス等
- クラスタ2:100-299床のへき地医療拠点、臨床研修指定、災害拠点の一般病院等
- クラスタ3:100-299床の一般病院等
- クラスタ4:300-499床の地域医療支援、臨床研修指定、災害拠点の一般病院等
- クラスタ5:300-499床の地域医療支援、へき地医療拠点、がん診療連携拠点、災害拠点の一般病院等
- クラスタ6:500床以上の特定機能、臨床研修指定、がん診療連携拠点、災害拠点の一般病院等
- クラスタ7:100-299床の臨床研修指定の一般病院等

In [4]:

```
#options(repos="http://cran.ism.ac.jp/")  
#install.packages("psych", dependencies=TRUE)  
library(psych)  
describeBy(z1, group=c1)
```

Descriptive statistics by group

group: 1

	vars	n	mean	sd	median	trimmed	mad	min	max	range	skew	kurtosis	se
q1	1	461	3.33	0.77	4	3.41	0	2	4	2	-0.63	-1.05	0.04
q2_1	2	461	2.00	0.00	2	2.00	0	2	2	0	NaN	NaN	0.00
q2_2	3	461	2.00	0.00	2	2.00	0	2	2	0	NaN	NaN	0.00
q2_3	4	461	2.00	0.00	2	2.00	0	2	2	0	NaN	NaN	0.00
q2_4	5	461	2.00	0.00	2	2.00	0	2	2	0	NaN	NaN	0.00
q2_5	6	461	2.00	0.00	2	2.00	0	2	2	0	NaN	NaN	0.00
q2_6	7	461	2.00	0.00	2	2.00	0	2	2	0	NaN	NaN	0.00
q3	8	461	3.00	0.57	3	2.99	0	1	5	4	0.34	2.63	0.03
c1	9	461	1.00	0.00	1	1.00	0	1	1	0	NaN	NaN	0.00

group: 2

	vars	n	mean	sd	median	trimmed	mad	min	max	range	skew	kurtosis	se
q1	1	87	1.69	1.25	1	1.51	0.00	1	4	3	1.26	-0.38	0.13
q2_1	2	87	2.00	0.00	2	2.00	0.00	2	2	0	NaN	NaN	0.00
q2_2	3	87	1.55	0.50	2	1.56	0.00	1	2	1	-0.20	-1.98	0.05
q2_3	4	87	1.00	0.00	1	1.00	0.00	1	1	0	NaN	NaN	0.00
q2_4	5	87	1.34	0.48	1	1.31	0.00	1	2	1	0.64	-1.61	0.05
q2_5	6	87	1.61	0.49	2	1.63	0.00	1	2	1	-0.44	-1.83	0.05
q2_6	7	87	1.44	0.50	1	1.42	0.00	1	2	1	0.25	-1.96	0.05
q3	8	87	3.51	0.94	3	3.54	1.48	1	5	4	-0.06	-0.21	0.10
c1	9	87	2.00	0.00	2	2.00	0.00	2	2	0	NaN	NaN	0.00

group: 3

	vars	n	mean	sd	median	trimmed	mad	min	max	range	skew	kurtosis	se
q1	1	434	1.23	0.45	1	1.15	0.00	1	3	2	1.65	1.66	0.02
q2_1	2	434	2.00	0.00	2	2.00	0.00	2	2	0	NaN	NaN	0.00
q2_2	3	434	2.00	0.00	2	2.00	0.00	2	2	0	NaN	NaN	0.00
q2_3	4	434	2.00	0.00	2	2.00	0.00	2	2	0	NaN	NaN	0.00
q2_4	5	434	2.00	0.00	2	2.00	0.00	2	2	0	NaN	NaN	0.00
q2_5	6	434	2.00	0.00	2	2.00	0.00	2	2	0	NaN	NaN	0.00
q2_6	7	434	2.00	0.00	2	2.00	0.00	2	2	0	NaN	NaN	0.00
q3	8	434	2.38	0.89	2	2.35	1.48	1	5	4	0.22	-0.23	0.04
c1	9	434	3.00	0.00	3	3.00	0.00	3	3	0	NaN	NaN	0.00

group: 4

	vars	n	mean	sd	median	trimmed	mad	min	max	range	skew	kurtosis	se
q1	1	185	1.46	1.04	1	1.21	0.00	1	4	3	1.85	1.58	0.08
q2_1	2	185	2.00	0.00	2	2.00	0.00	2	2	0	NaN	NaN	0.00
q2_2	3	185	1.25	0.43	1	1.19	0.00	1	2	1	1.15	-0.67	0.03
q2_3	4	185	2.00	0.00	2	2.00	0.00	2	2	0	NaN	NaN	0.00
q2_4	5	185	1.31	0.46	1	1.26	0.00	1	2	1	0.82	-1.33	0.03
q2_5	6	185	2.00	0.00	2	2.00	0.00	2	2	0	NaN	NaN	0.00
q2_6	7	185	1.42	0.50	1	1.40	0.00	1	2	1	0.31	-1.91	0.04
q3	8	185	3.62	0.88	4	3.65	1.48	1	5	4	-0.70	1.09	0.06
c1	9	185	4.00	0.00	4	4.00	0.00	4	4	0	NaN	NaN	0.00

group: 5

	vars	n	mean	sd	median	trimmed	mad	min	max	range	skew	kurtosis	se
q1	1	146	1.14	0.62	1	1.00	0.00	1	4	3	4.18	15.92	0.05
q2_1	2	146	2.00	0.00	2	2.00	0.00	2	2	0	NaN	NaN	0.00
q2_2	3	146	1.34	0.48	1	1.31	0.00	1	2	1	0.66	-1.58	0.04
q2_3	4	146	2.00	0.00	2	2.00	0.00	2	2	0	NaN	NaN	0.00
q2_4	5	146	1.07	0.25	1	1.00	0.00	1	2	1	3.38	9.50	0.02
q2_5	6	146	1.00	0.00	1	1.00	0.00	1	1	0	NaN	NaN	0.00
q2_6	7	146	1.35	0.48	1	1.31	0.00	1	2	1	0.63	-1.62	0.04
q3	8	146	4.34	0.76	4	4.46	1.48	2	5	3	-1.03	0.71	0.06
c1	9	146	5.00	0.00	5	5.00	0.00	5	5	0	NaN	NaN	0.00

group: 6

	vars	n	mean	sd	median	trimmed	mad	min	max	range	skew	kurtosis	se
q1	1	76	1.14	0.63	1	1.00	0	1	4	3	4.07	15.08	0.07
q2_1	2	76	1.00	0.00	1	1.00	0	1	1	0	NaN	NaN	0.00
q2_2	3	76	1.70	0.46	2	1.74	0	1	2	1	-0.84	-1.31	0.05
q2_3	4	76	1.91	0.29	2	2.00	0	1	2	1	-2.77	5.72	0.03
q2_4	5	76	1.13	0.34	1	1.05	0	1	2	1	2.14	2.60	0.04
q2_5	6	76	1.18	0.39	1	1.11	0	1	2	1	1.60	0.56	0.04
q2_6	7	76	1.25	0.44	1	1.19	0	1	2	1	1.13	-0.73	0.05
q3	8	76	4.67	0.72	5	4.84	0	1	5	4	-2.62	8.00	0.08
c1	9	76	6.00	0.00	6	6.00	0	6	6	0	NaN	NaN	0.00

group: 7

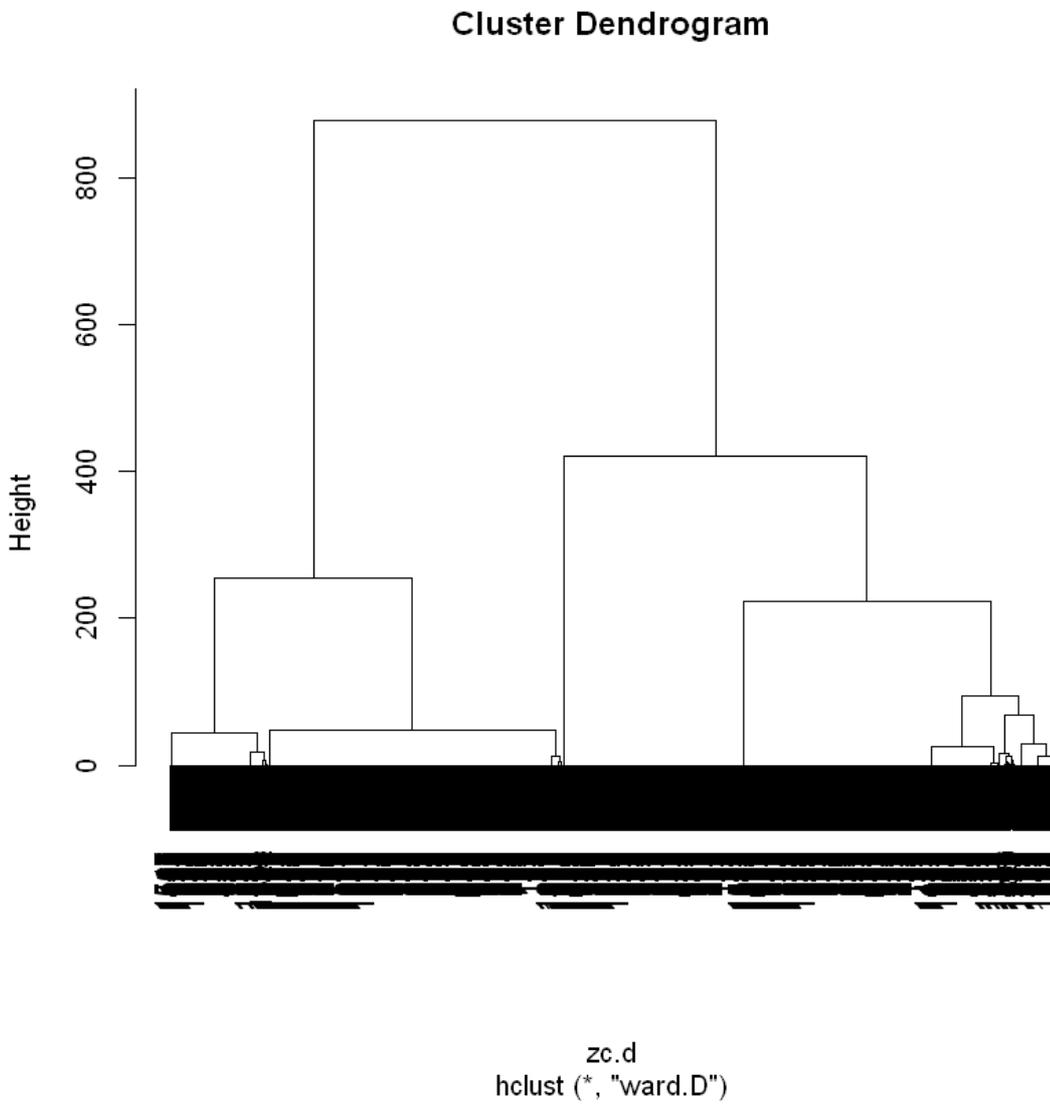
	vars	n	mean	sd	median	trimmed	mad	min	max	range	skew	kurtosis	se
q1	1	170	1.90	1.23	1	1.75	0	1	4	3	0.76	-1.20	0.09
q2_1	2	170	2.00	0.00	2	2.00	0	2	2	0	NaN	NaN	0.00
q2_2	3	170	2.00	0.00	2	2.00	0	2	2	0	NaN	NaN	0.00
q2_3	4	170	2.00	0.00	2	2.00	0	2	2	0	NaN	NaN	0.00
q2_4	5	170	1.00	0.00	1	1.00	0	1	1	0	NaN	NaN	0.00
q2_5	6	170	2.00	0.00	2	2.00	0	2	2	0	NaN	NaN	0.00
q2_6	7	170	2.00	0.00	2	2.00	0	2	2	0	NaN	NaN	0.00
q3	8	170	3.19	0.76	3	3.24	0	1	5	4	-0.34	0.43	0.06
c1	9	170	7.00	0.00	7	7.00	0	7	7	0	NaN	NaN	0.00

In [6]:

```
clusteredData<-merge(z_cluster,df, by="id")
```

In [7]:

```
zc<-clusteredData[,c("q105_1","q105_2","q105_3","q105_4","q105_5","q105_6")]  
idc <- clusteredData[,1:2]  
zc.d <- dist(zc)^2  
result_cluster_q105 <- hclust(zc.d, method="ward.D")  
plot(result_cluster_q105)
```



In [8]:

```
c_q105<-cutree(result_cluster_q105, k=9)
(z1c<-cbind(zc, c_q105))
```

q105_1	q105_2	q105_3	q105_4	q105_5	q105_6	c_q105
1	1	0	0	0	0	1
1	1	1	0	0	0	2
0	0	0	0	0	0	3
1	1	0	0	1	0	4
0	0	0	0	0	0	3
1	1	0	0	0	0	1
1	1	1	0	0	0	2
1	0	0	0	0	0	5
1	1	1	0	0	0	2
1	0	0	0	1	0	4
0	1	0	0	0	0	6
1	1	0	0	0	0	1
0	0	0	0	0	0	3
0	1	0	0	0	0	6
0	1	0	0	0	0	6
0	1	0	0	0	0	6
1	1	1	0	1	0	4
0	0	0	0	0	0	3
1	1	1	0	0	0	2
1	1	0	0	0	0	1
1	0	0	0	0	0	5
1	0	0	0	0	0	5
0	0	0	0	0	0	3
1	0	0	0	0	0	5
1	1	0	0	0	0	1
1	0	0	0	0	0	5
0	0	0	0	0	0	3
1	1	0	0	0	0	1
1	0	0	0	0	0	5
0	0	0	0	0	0	3
...
1	0	0	0	0	0	5
0	0	0	0	0	0	3
1	1	0	0	0	0	1

q105_1	q105_2	q105_3	q105_4	q105_5	q105_6	c_q105
1	0	0	0	0	0	5
1	0	0	0	0	0	5
1	1	0	0	0	0	1
0	0	0	0	0	0	3
0	0	0	0	0	0	3
1	0	0	0	0	0	5
1	1	0	0	0	0	1
1	1	1	0	1	0	4
1	1	0	0	1	0	4
0	0	0	0	0	0	3
1	1	0	0	0	0	1
1	1	0	0	0	0	1
1	1	0	0	0	0	1
1	1	0	0	0	0	1
0	1	0	0	0	0	6
1	1	0	0	0	0	1
1	1	0	0	0	0	1
0	0	0	0	0	0	3
1	0	0	0	0	0	5
0	0	0	0	0	0	3
1	1	0	0	0	0	1
1	0	0	0	0	0	5
0	0	0	0	0	0	3
1	1	0	0	0	0	1
1	1	1	0	1	0	4
1	1	0	0	1	0	4
1	1	0	0	0	0	1

薬剤部門にある医薬品に関するバーコードリーダー有無についてのクラスタの特徴を探るため、各クラスタにおける変数の平均値・分散について調べる。(結果のgroup:xは、クラスタxに相当) 結果より、それぞれのクラスタの特徴を鑑みて以下のように解釈することができる。

- クラスタ1:薬剤管理部門と調剤部門
- クラスタ2:薬剤管理部門と調剤部門と注射薬調剤部門
- クラスタ3:チェックなし
- クラスタ4:薬剤管理部門と調剤部門と注射薬調剤部門と無菌調剤部門
- クラスタ5:薬剤管理部門のみ
- クラスタ6:調剤部門のみ
- クラスタ7:薬剤管理部門と調剤部門と注射薬調剤部門と製剤部門

In [9]:

```
describeBy(z1c, group=c_q105)
```

Descriptive statistics by group

group: 1

	vars	n	mean	sd	median	trimmed	mad	min	max	range	skew	kurtosis	se
q105_1	1	330	1	0	1	1	0	1	1	0	NaN	NaN	0
q105_2	2	330	1	0	1	1	0	1	1	0	NaN	NaN	0
q105_3	3	330	0	0	0	0	0	0	0	0	NaN	NaN	0
q105_4	4	330	0	0	0	0	0	0	0	0	NaN	NaN	0
q105_5	5	330	0	0	0	0	0	0	0	0	NaN	NaN	0
q105_6	6	330	0	0	0	0	0	0	0	0	NaN	NaN	0
c_q105	7	330	1	0	1	1	0	1	1	0	NaN	NaN	0

group: 2

	vars	n	mean	sd	median	trimmed	mad	min	max	range	skew	kurtosis	se
q105_1	1	120	0.98	0.13	1	1.00	0	0	1	1	-7.46	54.05	0.01
q105_2	2	120	0.88	0.32	1	0.98	0	0	1	1	-2.36	3.59	0.03
q105_3	3	120	1.00	0.00	1	1.00	0	1	1	0	NaN	NaN	0.00
q105_4	4	120	0.00	0.00	0	0.00	0	0	0	0	NaN	NaN	0.00
q105_5	5	120	0.00	0.00	0	0.00	0	0	0	0	NaN	NaN	0.00
q105_6	6	120	0.00	0.00	0	0.00	0	0	0	0	NaN	NaN	0.00
c_q105	7	120	2.00	0.00	2	2.00	0	2	2	0	NaN	NaN	0.00

group: 3

	vars	n	mean	sd	median	trimmed	mad	min	max	range	skew	kurtosis	se
q105_1	1	498	0	0	0	0	0	0	0	0	NaN	NaN	0
q105_2	2	498	0	0	0	0	0	0	0	0	NaN	NaN	0
q105_3	3	498	0	0	0	0	0	0	0	0	NaN	NaN	0
q105_4	4	498	0	0	0	0	0	0	0	0	NaN	NaN	0
q105_5	5	498	0	0	0	0	0	0	0	0	NaN	NaN	0
q105_6	6	498	0	0	0	0	0	0	0	0	NaN	NaN	0
c_q105	7	498	3	0	3	3	0	3	3	0	NaN	NaN	0

group: 4

	vars	n	mean	sd	median	trimmed	mad	min	max	range	skew	kurtosis	se
q105_1	1	61	1.00	0.00	1	1.00	0	1	1	0	NaN	NaN	0.00
q105_2	2	61	0.89	0.32	1	0.98	0	0	1	1	-2.36	3.62	0.04
q105_3	3	61	0.51	0.50	1	0.51	0	0	1	1	-0.03	-2.03	0.06
q105_4	4	61	0.00	0.00	0	0.00	0	0	0	0	NaN	NaN	0.00
q105_5	5	61	1.00	0.00	1	1.00	0	1	1	0	NaN	NaN	0.00
q105_6	6	61	0.00	0.00	0	0.00	0	0	0	0	NaN	NaN	0.00
c_q105	7	61	4.00	0.00	4	4.00	0	4	4	0	NaN	NaN	0.00

group: 5

	vars	n	mean	sd	median	trimmed	mad	min	max	range	skew	kurtosis	se
q105_1	1	317	1	0	1	1	0	1	1	0	NaN	NaN	0
q105_2	2	317	0	0	0	0	0	0	0	0	NaN	NaN	0
q105_3	3	317	0	0	0	0	0	0	0	0	NaN	NaN	0
q105_4	4	317	0	0	0	0	0	0	0	0	NaN	NaN	0
q105_5	5	317	0	0	0	0	0	0	0	0	NaN	NaN	0
q105_6	6	317	0	0	0	0	0	0	0	0	NaN	NaN	0
c_q105	7	317	5	0	5	5	0	5	5	0	NaN	NaN	0

group: 6

	vars	n	mean	sd	median	trimmed	mad	min	max	range	skew	kurtosis	se
q105_1	1	140	0	0	0	0	0	0	0	0	NaN	NaN	0
q105_2	2	140	1	0	1	1	0	1	1	0	NaN	NaN	0
q105_3	3	140	0	0	0	0	0	0	0	0	NaN	NaN	0
q105_4	4	140	0	0	0	0	0	0	0	0	NaN	NaN	0
q105_5	5	140	0	0	0	0	0	0	0	0	NaN	NaN	0
q105_6	6	140	0	0	0	0	0	0	0	0	NaN	NaN	0
c_q105	7	140	6	0	6	6	0	6	6	0	NaN	NaN	0

group: 7

	vars	n	mean	sd	median	trimmed	mad	min	max	range	skew	kurtosis	se
q105_1	1	40	0.98	0.16	1	1.00	0	0	1	1	-5.86	33.15	0.02
q105_2	2	40	0.92	0.27	1	1.00	0	0	1	1	-3.11	7.85	0.04
q105_3	3	40	0.78	0.42	1	0.84	0	0	1	1	-1.27	-0.40	0.07
q105_4	4	40	0.98	0.16	1	1.00	0	0	1	1	-5.86	33.15	0.02
q105_5	5	40	0.38	0.49	0	0.34	0	0	1	1	0.50	-1.80	0.08
q105_6	6	40	0.08	0.27	0	0.00	0	0	1	1	3.11	7.85	0.04
c_q105	7	40	7.00	0.00	7	7.00	0	7	7	0	NaN	NaN	0.00

group: 8

	vars	n	mean	sd	median	trimmed	mad	min	max	range	skew	kurtosis	se
q105_1	1	32	0.00	0.00	0	0.00	0	0	0	0	NaN	NaN	0.00
q105_2	2	32	0.97	0.18	1	1.00	0	0	1	1	-5.14	25.18	0.03
q105_3	3	32	0.84	0.37	1	0.92	0	0	1	1	-1.81	1.30	0.07
q105_4	4	32	0.00	0.00	0	0.00	0	0	0	0	NaN	NaN	0.00
q105_5	5	32	0.38	0.49	0	0.35	0	0	1	1	0.49	-1.81	0.09
q105_6	6	32	0.03	0.18	0	0.00	0	0	1	1	5.14	25.18	0.03
c_q105	7	32	8.00	0.00	8	8.00	0	8	8	0	NaN	NaN	0.00

group: 9

	vars	n	mean	sd	median	trimmed	mad	min	max	range	skew	kurtosis	se
q105_1	1	21	0.14	0.36	0	0.06	0	0	1	1	1.90	1.69	0.08
q105_2	2	21	0.43	0.51	0	0.41	0	0	1	1	0.27	-2.02	0.11
q105_3	3	21	0.00	0.00	0	0.00	0	0	0	0	NaN	NaN	0.00
q105_4	4	21	0.00	0.00	0	0.00	0	0	0	0	NaN	NaN	0.00
q105_5	5	21	0.00	0.00	0	0.00	0	0	0	0	NaN	NaN	0.00
q105_6	6	21	1.00	0.00	1	1.00	0	1	1	0	NaN	NaN	0.00
c_q105	7	21	9.00	0.00	9	9.00	0	9	9	0	NaN	NaN	0.00

In [10]:

```
z2c<-cbind(idc, z1c)
z3c<-z2c[, c("id", "c_q105")]
clusteredData2<-merge(z3c, clusteredData, by="id")
z_final<-clusteredData2[clusteredData2$q106_1>0, ]
z_final$q106_1[z_final$q106_1==2]<-1
```

In [17]:

```
z_final_for_cross=z_final[, c("c1", "c_q105")]
xtabs(~c1+c_q105, z_final_for_cross)
```

```
  c_q105
c1    1    2    3    4    5    6    7    8    9
  1  81  14 185    1 107  50    4    2    8
  2  23  11  13    5  15    7    3    4    3
  3  61  19 205    2 100  33    1    5    5
  4  49  28  32   14  34   12    5    4    3
  5  47  23  10   19  11   13   12    8    0
  6  14  10    5   13    9    6   11    6    1
  7  47  13  40    5   39   18    3    2    1
```

In [14]:

```
#options(repos="http://cran.ism.ac.jp/")
#install.packages("partykit", dependencies=TRUE)
#install.packages("rpart.plot", dependencies=TRUE)
```

In [13]:

```
library(rpart)
library(rpart.plot)
library(partykit)
```

目的変数を「バーコードシステムを導入あるいは拡大する計画はありますか」（1:ある、2:検討中である、3:計画はない）の1と2を1にまとめたものとし、説明変数を「薬剤部門にある医薬品に関するバーコードリーダー有無」のクラスタ、医薬品に関するバーコードの利用有無およびq1からq3とした。

In [14]:

```
ct <- rpart(q106_1~q1+q2_1+q2_2+q2_3+q2_4+q2_5+q2_6+q3+q101+c_q105, data = z_final, method='class')
print(ct)
```

n= 1534

```
node), split, n, loss, yval, (yprob)
* denotes terminal node
```

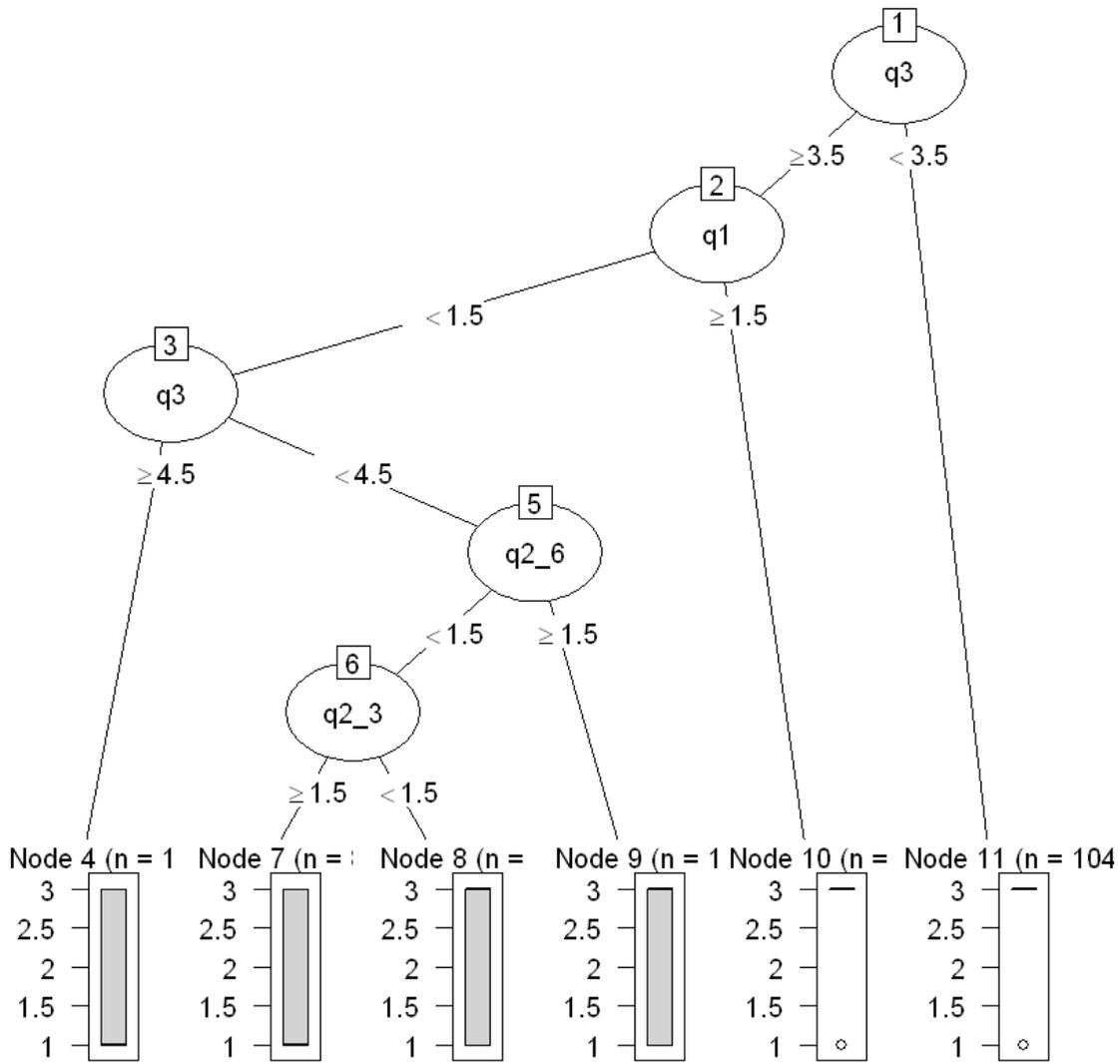
```
1) root 1534 405 3 (0.2640156 0.7359844)
 2) q3>=3.5 488 219 3 (0.4487705 0.5512295)
   4) q1< 1.5 370 180 1 (0.5135135 0.4864865)
     8) q3>=4.5 163 67 1 (0.5889571 0.4110429) *
     9) q3< 4.5 207 94 3 (0.4541063 0.5458937)
       18) q2_6< 1.5 97 45 1 (0.5360825 0.4639175)
         36) q2_3>=1.5 80 34 1 (0.5750000 0.4250000) *
         37) q2_3< 1.5 17 6 3 (0.3529412 0.6470588) *
       19) q2_6>=1.5 110 42 3 (0.3818182 0.6181818) *
   5) q1>=1.5 118 29 3 (0.2457627 0.7542373) *
 3) q3< 3.5 1046 186 3 (0.1778203 0.8221797) *
```

- 300床未満の病院では導入の計画はないという回答が多い傾向があり、
- 300床以上の大きな病院においては
 - 一般病院でない場合には導入の計画はないという回答が多い傾向があり、
 - 一般病院の場合、
 - 500床以上であれば計画はあり、
 - 500床未満では
 - 災害拠点病院であれば計画はあり
 - 災害拠点病院でなければ計画はないという回答が多い傾向が見られた

このことから、病院の規模が導入・拡張有無に大きく影響があることが見て取れる

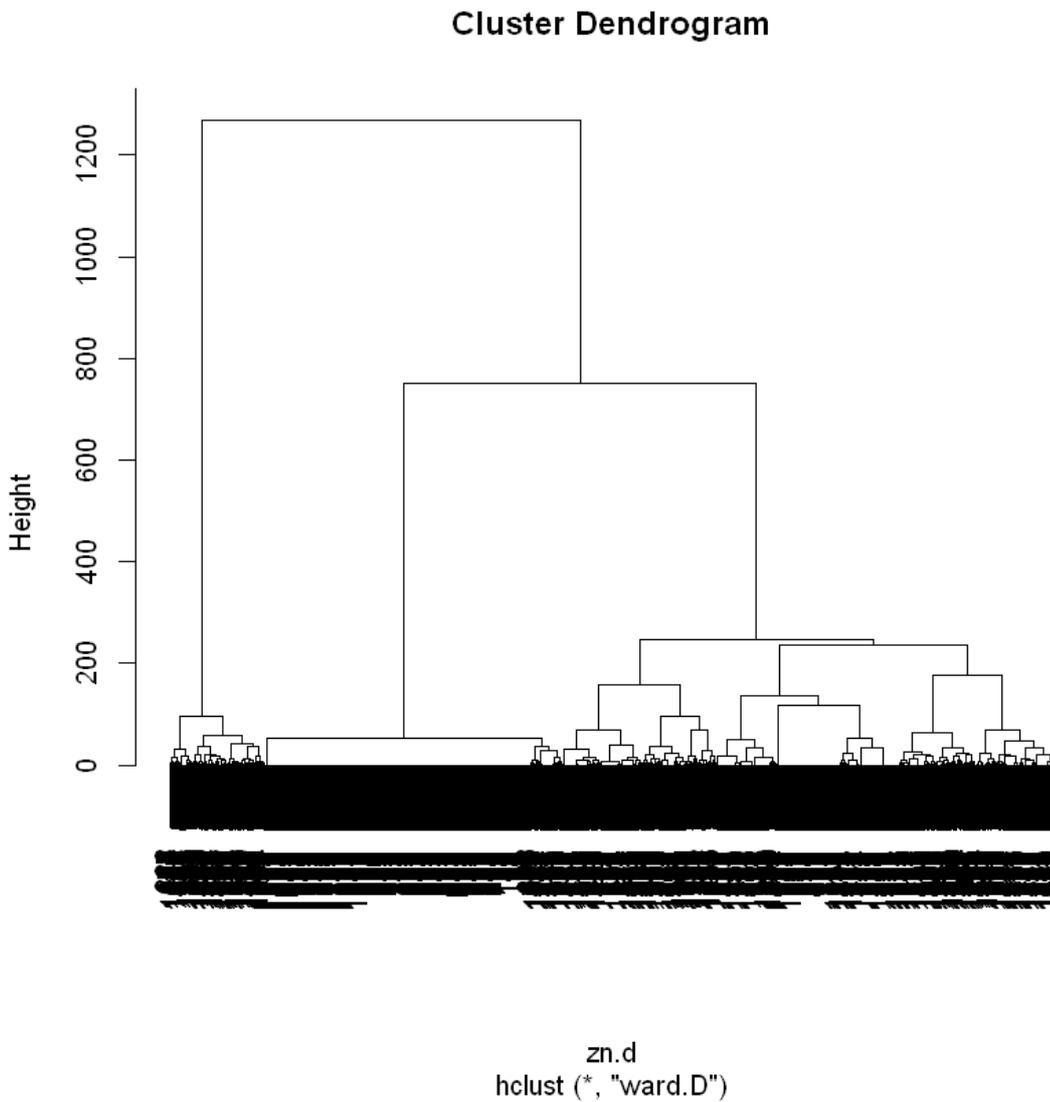
In [15]:

```
plot(as.party(ct))
```



In [22]:

```
z_needs<-z_final[, c("q301_1", "q301_2", "q301_3", "q301_4", "q301_5", "q301_6", "q301_7", "q301_8", "q301_9", "q301_10", "q301_11", "q301_12", "q301_13", "q301_14")]
id<-z_final[, 1]
zn.d <- dist(z_needs)^2
result_needs_cluster <- hclust(zn.d, method="ward.D")
plot(result_needs_cluster)
```

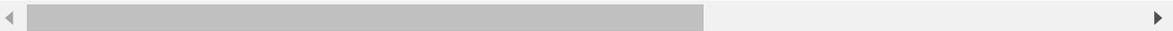


In [23]:

```
c_needs<-cutree(result_needs_cluster, k=5)
(z1needs<-cbind(z_needs, c_needs))
```

	q301_1	q301_2	q301_3	q301_4	q301_5	q301_6	q301_7	q301_8	q301_9	q3
1	0	0	0	0	0	0	0	0	0	0
2	1	1	0	0	0	0	0	0	0	0
3	0	0	0	0	0	0	0	0	0	0
4	1	1	0	0	0	0	0	0	0	0
5	0	0	0	0	0	0	0	0	0	0
6	1	1	1	1	1	1	1	1	1	0
7	0	1	1	0	0	0	0	1	1	0
8	0	1	0	0	0	0	0	0	0	0
9	0	0	0	0	0	0	0	0	0	0
10	0	1	0	1	0	0	0	1	0	0
11	0	0	0	0	0	0	0	0	0	0
12	0	0	0	0	0	0	0	0	0	0
13	0	0	0	0	0	0	0	0	0	0
14	0	1	0	0	0	0	0	0	0	0
15	0	1	0	1	1	1	1	0	1	0
16	0	0	0	0	0	0	0	0	0	0
17	0	1	0	0	0	0	0	0	0	0
18	0	0	0	0	0	0	0	0	0	0
19	0	1	1	0	0	0	0	1	1	0
20	0	1	0	0	0	0	0	0	0	0
21	0	1	0	0	0	0	1	0	0	0
22	0	0	0	0	0	0	0	0	0	0
23	0	0	0	0	0	0	0	0	0	0
24	0	1	0	0	0	0	0	0	0	0
25	0	0	0	0	0	0	0	0	0	0
26	0	0	0	0	0	0	0	0	0	0
27	0	1	0	0	0	0	0	0	0	0
28	1	1	0	0	0	0	0	0	0	0
29	0	1	0	0	1	0	0	0	0	0
30	1	0	0	0	0	0	1	0	0	0
...
1530	0	0	0	0	0	0	0	0	0	0
1531	0	0	0	0	0	0	0	0	0	0
1532	0	1	0	0	0	0	0	0	0	0

	q301_1	q301_2	q301_3	q301_4	q301_5	q301_6	q301_7	q301_8	q301_9	q3
1533	0	0	0	0	0	0	0	0	0	0
1534	0	0	0	0	0	0	0	0	0	0
1535	1	0	1	0	0	0	0	0	1	0
1536	1	1	0	0	0	0	1	0	0	0
1537	0	0	0	0	0	0	0	0	0	0
1538	0	0	0	0	0	0	0	0	0	0
1539	0	0	0	0	0	0	0	0	0	0
1540	1	1	0	1	0	0	1	0	0	0
1541	0	0	0	0	0	0	0	0	0	0
1542	0	0	0	0	0	0	0	0	0	0
1543	1	1	0	1	1	1	0	0	0	1
1544	0	0	0	0	0	0	0	0	0	0
1545	0	1	0	0	0	0	0	0	0	0
1546	0	1	0	0	0	0	1	1	0	0
1547	0	0	0	0	0	1	1	0	0	0
1548	1	1	1	0	1	0	0	0	0	0
1549	0	1	0	0	0	0	0	0	0	0
1550	0	0	0	0	0	0	0	0	0	0
1551	0	0	0	1	1	1	1	0	0	0
1552	0	0	0	0	0	0	0	0	0	0
1553	0	1	0	0	1	1	1	0	0	0
1554	0	0	0	0	0	0	0	0	0	0
1555	0	0	0	0	0	0	0	0	0	0
1556	1	1	0	0	0	0	1	1	1	0
1557	1	1	0	0	0	0	0	0	0	0
1558	0	0	0	0	0	0	0	0	0	0
1559	0	0	0	0	0	0	0	0	0	0



バーコードを変更してほしい事例についての5つのクラスタについては以下のように解釈できる。

- クラスタ1:回答なし
- クラスタ2:主に剤形・有効成分に関する情報を付した販売名への変更、ブランド名の変更、販売会社変更時、銘柄別⇔統一名収載変更時
- クラスタ3:主に剤形・有効成分に関する情報を付した販売名への変更、ブランド名の変更、製剤の色・形・大きさ・識別記号の変更、販売会社変更時、管理保管方法変更時、銘柄別⇔統一名収載変更時
- クラスタ4:主にブランド名の変更のみ
- クラスタ5:主に有効成分以外の成分・分量変更時、管理保管方法変更時

In [24]:

```
describeBy(z_needs, group=c_needs)
```

Descriptive statistics by group

group: 1

	vars	n	mean	sd	median	trimmed	mad	min	max	range	skew	kurtosis	se
q301_1	1	515	0.06	0.23	0	0	0	0	1	1	3.84	12.76	0.01
q301_2	2	515	0.02	0.14	0	0	0	0	1	1	6.60	41.67	0.01
q301_3	3	515	0.01	0.10	0	0	0	0	1	1	9.97	97.62	0.00
q301_4	4	515	0.01	0.10	0	0	0	0	1	1	9.97	97.62	0.00
q301_5	5	515	0.00	0.04	0	0	0	0	1	1	22.56	508.01	0.00
q301_6	6	515	0.00	0.00	0	0	0	0	0	0	NaN	NaN	0.00
q301_7	7	515	0.03	0.17	0	0	0	0	1	1	5.39	27.10	0.01
q301_8	8	515	0.02	0.13	0	0	0	0	1	1	7.34	52.03	0.01
q301_9	9	515	0.03	0.17	0	0	0	0	1	1	5.58	29.24	0.01
q301_10	10	515	0.00	0.06	0	0	0	0	1	1	15.91	251.51	0.00
q301_11	11	515	0.01	0.08	0	0	0	0	1	1	12.95	166.01	0.00
q301_12	12	515	0.00	0.06	0	0	0	0	1	1	15.91	251.51	0.00
q301_13	13	515	0.00	0.00	0	0	0	0	0	0	NaN	NaN	0.00
q301_14	14	515	0.00	0.06	0	0	0	0	1	1	15.91	251.51	0.00

group: 2

	vars	n	mean	sd	median	trimmed	mad	min	max	range	skew	kurtosis	se
q301_1	1	267	0.72	0.45	1	0.78	0	0	1	1	-0.99	-1.02	0.03
q301_2	2	267	0.94	0.24	1	1.00	0	0	1	1	-3.69	11.64	0.01
q301_3	3	267	0.13	0.34	0	0.04	0	0	1	1	2.17	2.74	0.02
q301_4	4	267	0.05	0.22	0	0.00	0	0	1	1	4.17	15.45	0.01
q301_5	5	267	0.14	0.35	0	0.06	0	0	1	1	2.04	2.15	0.02
q301_6	6	267	0.03	0.17	0	0.00	0	0	1	1	5.48	28.17	0.01
q301_7	7	267	0.23	0.42	0	0.17	0	0	1	1	1.26	-0.41	0.03
q301_8	8	267	0.11	0.31	0	0.01	0	0	1	1	2.50	4.27	0.02
q301_9	9	267	0.09	0.29	0	0.00	0	0	1	1	2.77	5.72	0.02
q301_10	10	267	0.06	0.24	0	0.00	0	0	1	1	3.69	11.64	0.01
q301_11	11	267	0.64	0.48	1	0.68	0	0	1	1	-0.60	-1.65	0.03
q301_12	12	267	0.42	0.49	0	0.40	0	0	1	1	0.32	-1.90	0.03
q301_13	13	267	0.20	0.40	0	0.13	0	0	1	1	1.50	0.26	0.02
q301_14	14	267	0.58	0.49	1	0.60	0	0	1	1	-0.34	-1.89	0.03

group: 3

	vars	n	mean	sd	median	trimmed	mad	min	max	range	skew	kurtosis	se
q301_1	1	165	0.52	0.50	1	0.53	0	0	1	1	-0.08	-2.00	0.04
q301_2	2	165	0.89	0.31	1	0.98	0	0	1	1	-2.49	4.20	0.02
q301_3	3	165	0.49	0.50	0	0.49	0	0	1	1	0.04	-2.01	0.04
q301_4	4	165	0.95	0.23	1	1.00	0	0	1	1	-3.89	13.19	0.02
q301_5	5	165	0.97	0.17	1	1.00	0	0	1	1	-5.43	27.66	0.01
q301_6	6	165	0.92	0.28	1	1.00	0	0	1	1	-2.95	6.76	0.02
q301_7	7	165	0.90	0.30	1	0.99	0	0	1	1	-2.59	4.73	0.02
q301_8	8	165	0.37	0.48	0	0.34	0	0	1	1	0.53	-1.72	0.04
q301_9	9	165	0.20	0.40	0	0.13	0	0	1	1	1.49	0.21	0.03
q301_10	10	165	0.25	0.44	0	0.20	0	0	1	1	1.12	-0.76	0.03
q301_11	11	165	0.53	0.50	1	0.53	0	0	1	1	-0.11	-2.00	0.04
q301_12	12	165	0.75	0.43	1	0.81	0	0	1	1	-1.15	-0.67	0.03
q301_13	13	165	0.44	0.50	0	0.43	0	0	1	1	0.23	-1.96	0.04
q301_14	14	165	0.55	0.50	1	0.56	0	0	1	1	-0.21	-1.97	0.04

group: 4

	vars	n	mean	sd	median	trimmed	mad	min	max	range	skew	kurtosis	se
q301_1	1	316	0.00	0.06	0	0.00	0	0	1	1	17.61	309.02	0.00
q301_2	2	316	0.83	0.37	1	0.91	0	0	1	1	-1.77	1.14	0.02
q301_3	3	316	0.02	0.14	0	0.00	0	0	1	1	7.02	47.37	0.01
q301_4	4	316	0.01	0.10	0	0.00	0	0	1	1	10.07	99.69	0.01
q301_5	5	316	0.00	0.00	0	0.00	0	0	0	0	NaN	NaN	0.00
q301_6	6	316	0.00	0.00	0	0.00	0	0	0	0	NaN	NaN	0.00
q301_7	7	316	0.09	0.28	0	0.00	0	0	1	1	2.95	6.74	0.02

q301_8	8	316	0.02	0.15	0	0.00	0	0	1	1	6.46	39.89	0.01
q301_9	9	316	0.00	0.00	0	0.00	0	0	0	0	NaN	NaN	0.00
q301_10	10	316	0.01	0.10	0	0.00	0	0	1	1	10.07	99.69	0.01
q301_11	11	316	0.46	0.50	0	0.44	0	0	1	1	0.18	-1.97	0.03
q301_12	12	316	0.10	0.31	0	0.01	0	0	1	1	2.57	4.64	0.02
q301_13	13	316	0.05	0.21	0	0.00	0	0	1	1	4.24	16.00	0.01
q301_14	14	316	0.32	0.47	0	0.27	0	0	1	1	0.79	-1.39	0.03

group: 5

	vars	n	mean	sd	median	trimmed	mad	min	max	range	skew	kurtosis	se
q301_1	1	271	0.25	0.43	0	0.19	0	0	1	1	1.14	-0.70	0.03
q301_2	2	271	0.49	0.50	0	0.49	0	0	1	1	0.04	-2.01	0.03
q301_3	3	271	0.51	0.50	1	0.51	0	0	1	1	-0.02	-2.01	0.03
q301_4	4	271	0.03	0.17	0	0.00	0	0	1	1	5.53	28.67	0.01
q301_5	5	271	0.02	0.13	0	0.00	0	0	1	1	7.12	48.83	0.01
q301_6	6	271	0.01	0.12	0	0.00	0	0	1	1	8.00	62.28	0.01
q301_7	7	271	0.11	0.31	0	0.01	0	0	1	1	2.53	4.41	0.02
q301_8	8	271	0.03	0.17	0	0.00	0	0	1	1	5.53	28.67	0.01
q301_9	9	271	0.03	0.18	0	0.00	0	0	1	1	5.18	24.94	0.01
q301_10	10	271	0.06	0.24	0	0.00	0	0	1	1	3.59	10.90	0.01
q301_11	11	271	0.25	0.43	0	0.18	0	0	1	1	1.17	-0.64	0.03
q301_12	12	271	0.64	0.48	1	0.68	0	0	1	1	-0.59	-1.66	0.03
q301_13	13	271	0.27	0.45	0	0.22	0	0	1	1	1.01	-0.98	0.03
q301_14	14	271	0.28	0.45	0	0.23	0	0	1	1	0.97	-1.06	0.03

In [25]:

```
z_needs<-cbind(z_final,c_needs)
z_needs_final<-z_needs[z_needs$q204>0 & z_needs$q106_1>0 & z_needs$q1>0 & z_needs$q2_1>0 & z_needs$q2_2>0 & z_needs$q2_3>0 & z_needs$q2_4>0 & z_needs$q2_5>0 & z_needs$q2_6>0 & z_needs$q3>0 & z_needs$q101>0 & z_needs$c_q105>0,]
```

In [26]:

```
ct3 <- rpart(c_needs~q204+q106_1+q1+q2_1+q2_2+q2_3+q2_4+q2_5+q2_6+q3+q101+c_q105, data = z_needs_final, method='class')
print(ct3)
```

n= 1016

node), split, n, loss, yval, (yprob)
* denotes terminal node

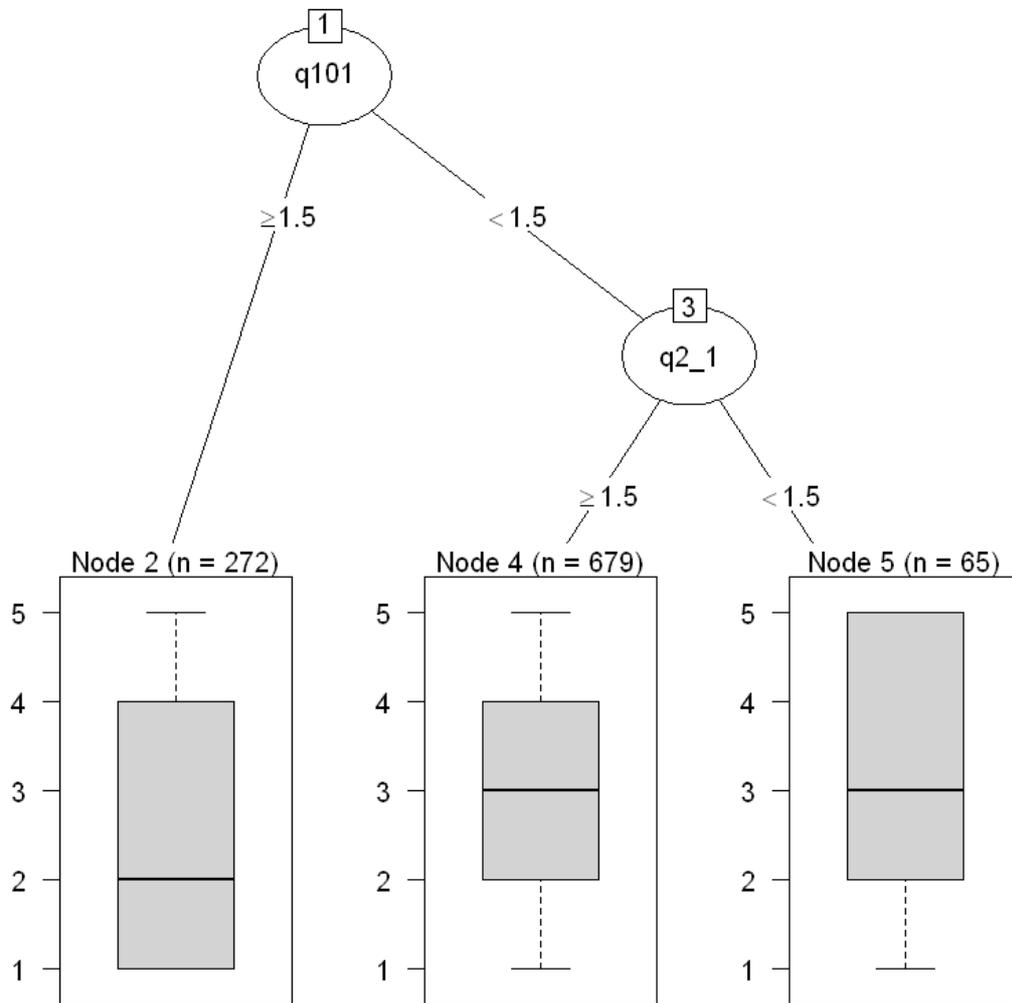
- 1) root 1016 726 1 (0.29 0.19 0.1 0.22 0.19)
- 2) q101>=1.5 272 138 1 (0.49 0.12 0.096 0.13 0.16) *
- 3) q101< 1.5 744 553 4 (0.21 0.22 0.11 0.26 0.21)
- 6) q2_1>=1.5 679 498 4 (0.22 0.2 0.11 0.27 0.2) *
- 7) q2_1< 1.5 65 41 2 (0.077 0.37 0.12 0.15 0.28) *

決定木分析の結果より、以下のように回答されている傾向にある。

- バーコードを利用していない場合、クラス1（回答なし）
- バーコードを利用している場合、
 - 特定機能病院でない場合、クラス4（主にブランド名の変更のみ）
 - 特定機能病院である場合、クラス2（主に剤形・有効成分に関する情報を付した販売名への変更、ブランド名の変更、販売会社変更時、銘柄別⇔統一名収載変更時）

In [97]:

```
plot(as.party(ct3))
```



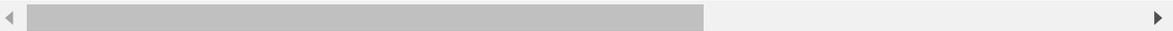
以下は、バーコードを変更してほしい事例についてのクラスター別（クラスター1、2、4のみ）の実データ例を示す。

In [100]:

```
z1needs[z1needs$c_needs==1,]
```

	q301_1	q301_2	q301_3	q301_4	q301_5	q301_6	q301_7	q301_8	q301_9	q3
1	0	0	0	0	0	0	0	0	0	0
3	0	0	0	0	0	0	0	0	0	0
5	0	0	0	0	0	0	0	0	0	0
9	0	0	0	0	0	0	0	0	0	0
11	0	0	0	0	0	0	0	0	0	0
13	0	0	0	0	0	0	0	0	0	0
16	0	0	0	0	0	0	0	0	0	0
18	0	0	0	0	0	0	0	0	0	0
23	0	0	0	0	0	0	0	0	0	0
25	0	0	0	0	0	0	0	0	0	0
26	0	0	0	0	0	0	0	0	0	0
33	0	0	0	0	0	0	0	0	0	0
36	0	0	0	0	0	0	0	0	0	0
40	1	0	0	0	0	0	0	0	0	0
41	0	1	1	1	0	0	1	0	0	0
52	0	0	0	0	0	0	0	0	0	0
54	0	0	0	0	0	0	0	0	0	0
60	0	0	0	0	0	0	0	0	0	0
61	0	0	0	0	0	0	0	0	0	0
62	0	0	0	0	0	0	0	0	0	0
73	0	0	0	0	0	0	0	0	0	0
75	0	0	0	0	0	0	0	0	0	0
76	0	0	0	0	0	0	0	0	0	0
77	0	0	0	0	0	0	0	0	0	0
83	0	0	0	0	0	0	0	0	0	0
84	0	0	0	0	0	0	0	0	0	0
85	1	0	0	0	0	0	0	0	1	0
86	0	0	0	0	0	0	0	0	0	0
87	0	0	0	0	0	0	0	0	0	0
92	0	0	0	0	0	0	0	0	0	0
...
1481	0	0	0	0	0	0	0	0	0	0
1483	0	0	0	0	0	0	0	0	0	0
1484	0	0	0	0	0	0	0	0	0	0

	q301_1	q301_2	q301_3	q301_4	q301_5	q301_6	q301_7	q301_8	q301_9	q3
1487	0	0	0	0	0	0	0	0	0	0
1488	0	0	0	0	0	0	0	0	0	0
1491	0	0	0	0	0	0	0	0	0	0
1493	0	0	0	0	0	0	0	0	0	0
1496	0	0	0	0	0	0	0	0	0	0
1497	0	0	0	0	0	0	0	0	0	0
1499	0	0	0	0	0	0	0	0	0	0
1500	0	0	0	0	0	0	0	0	0	0
1506	0	0	0	0	0	0	0	0	0	0
1508	0	0	0	0	0	0	0	0	0	0
1509	0	0	0	0	0	0	0	0	0	0
1517	0	0	0	0	0	0	0	0	0	0
1520	0	0	0	0	0	0	0	0	0	0
1521	0	0	0	0	0	0	0	0	0	0
1525	0	0	0	0	0	0	0	0	0	0
1529	0	0	0	0	0	0	0	0	0	0
1530	0	0	0	0	0	0	0	0	0	0
1531	0	0	0	0	0	0	0	0	0	0
1533	0	0	0	0	0	0	0	0	0	0
1534	0	0	0	0	0	0	0	0	0	0
1537	0	0	0	0	0	0	0	0	0	0
1541	0	0	0	0	0	0	0	0	0	0
1552	0	0	0	0	0	0	0	0	0	0
1554	0	0	0	0	0	0	0	0	0	0
1555	0	0	0	0	0	0	0	0	0	0
1558	0	0	0	0	0	0	0	0	0	0
1559	0	0	0	0	0	0	0	0	0	0



In [27]:

```
z1needs[z1needs$c_needs==2, ]
```

	q301_1	q301_2	q301_3	q301_4	q301_5	q301_6	q301_7	q301_8	q301_9	q3
2	1	1	0	0	0	0	0	0	0	0
4	1	1	0	0	0	0	0	0	0	0
7	0	1	1	0	0	0	0	1	1	0
28	1	1	0	0	0	0	0	0	0	0
29	0	1	0	0	1	0	0	0	0	0
31	0	1	0	0	0	0	0	1	1	0
37	1	1	0	0	0	0	0	0	0	0
44	1	1	1	0	0	0	0	0	0	0
45	1	1	0	0	0	0	0	0	0	0
46	1	1	0	0	0	0	0	0	0	0
47	0	1	0	0	0	0	0	0	0	0
48	0	1	0	0	0	0	0	0	0	0
50	1	1	0	0	0	0	0	0	0	0
51	1	1	1	0	0	0	1	0	0	0
56	1	1	1	0	0	0	1	0	1	0
58	1	1	0	0	0	0	0	0	0	0
64	1	1	1	0	0	0	0	0	0	1
71	1	1	0	0	0	0	0	0	0	0
72	0	1	0	0	0	0	0	0	0	0
89	0	1	0	0	0	0	1	0	1	1
102	1	1	0	0	0	0	0	0	0	0
117	1	0	0	0	0	0	0	1	0	0
122	1	1	0	0	0	0	0	0	0	0
135	0	1	0	0	0	0	0	0	0	0
137	1	1	0	0	0	0	0	0	0	0
138	1	1	0	0	0	0	0	0	0	0
141	1	1	0	0	0	0	0	0	0	0
144	0	1	0	0	1	0	1	0	0	0
154	1	0	0	0	0	0	0	0	0	0
161	1	1	0	1	1	0	0	0	0	0
...
1365	1	1	0	0	0	0	0	0	0	0
1368	1	1	0	0	0	0	0	0	0	0
1369	0	1	0	0	0	0	1	0	0	0

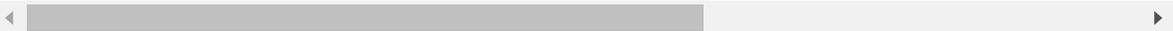
	q301_1	q301_2	q301_3	q301_4	q301_5	q301_6	q301_7	q301_8	q301_9	q3
1380	1	1	0	0	0	0	0	0	0	0
1385	1	1	0	0	0	0	0	0	0	0
1387	0	1	0	0	1	0	0	0	0	0
1405	0	1	0	0	0	0	0	0	0	0
1419	1	0	0	0	0	0	0	0	0	0
1421	0	1	0	0	0	0	0	1	0	0
1427	1	1	0	0	0	0	1	0	0	0
1428	1	1	0	0	0	0	0	0	0	0
1429	0	1	0	1	1	1	0	0	0	0
1431	1	1	0	0	0	0	0	0	0	0
1446	1	1	0	0	0	0	0	1	0	0
1453	1	1	0	0	1	0	0	0	0	0
1462	0	1	0	0	0	0	0	0	0	0
1469	0	1	1	0	0	0	0	1	1	0
1471	0	1	0	0	0	0	0	0	0	0
1472	1	1	0	0	0	0	0	0	0	0
1490	1	1	0	0	0	0	1	1	1	0
1501	1	1	1	0	0	0	1	0	0	0
1513	1	1	0	0	0	0	0	0	0	0
1522	1	1	0	0	0	0	0	0	0	0
1526	1	1	0	0	0	0	0	0	0	0
1527	0	1	0	0	1	0	1	0	0	0
1536	1	1	0	0	0	0	1	0	0	0
1540	1	1	0	1	0	0	1	0	0	0
1548	1	1	1	0	1	0	0	0	0	0
1556	1	1	0	0	0	0	1	1	1	0
1557	1	1	0	0	0	0	0	0	0	0

In [28]:

```
z1needs[z1needs$c_needs==4, ]
```

	q301_1	q301_2	q301_3	q301_4	q301_5	q301_6	q301_7	q301_8	q301_9	q3
8	0	1	0	0	0	0	0	0	0	0
10	0	1	0	1	0	0	0	1	0	0
14	0	1	0	0	0	0	0	0	0	0
17	0	1	0	0	0	0	0	0	0	0
20	0	1	0	0	0	0	0	0	0	0
21	0	1	0	0	0	0	1	0	0	0
22	0	0	0	0	0	0	0	0	0	0
24	0	1	0	0	0	0	0	0	0	0
27	0	1	0	0	0	0	0	0	0	0
38	0	1	0	0	0	0	0	0	0	0
43	0	1	0	0	0	0	0	0	0	0
53	0	1	0	0	0	0	0	0	0	0
55	0	1	0	0	0	0	1	0	0	0
57	0	1	0	0	0	0	1	0	0	1
63	0	1	0	0	0	0	0	0	0	0
70	0	1	0	0	0	0	0	0	0	0
78	0	1	0	0	0	0	0	0	0	0
80	0	1	0	0	0	0	0	0	0	0
81	0	1	0	0	0	0	0	0	0	0
88	0	1	0	0	0	0	0	0	0	0
98	0	1	0	0	0	0	1	0	0	0
99	0	1	0	0	0	0	1	0	0	0
100	0	1	0	1	0	0	0	0	0	0
101	0	1	0	0	0	0	0	0	0	0
103	0	0	0	0	0	0	0	0	0	0
104	0	0	0	0	0	0	0	0	0	0
109	0	1	0	0	0	0	0	0	0	0
111	0	0	0	0	0	0	0	0	0	0
113	0	1	0	0	0	0	0	0	0	0
114	0	1	1	0	0	0	0	0	0	0
...
1415	0	1	0	0	0	0	0	0	0	0
1426	0	1	0	0	0	0	0	0	0	0
1430	0	0	0	0	0	0	0	0	0	0

	q301_1	q301_2	q301_3	q301_4	q301_5	q301_6	q301_7	q301_8	q301_9	q3
1432	0	1	0	0	0	0	1	1	0	0
1439	0	0	0	0	0	0	0	0	0	0
1443	0	0	0	0	0	0	0	0	0	0
1449	0	1	0	0	0	0	1	0	0	0
1451	0	1	0	0	0	0	0	0	0	0
1463	0	1	0	0	0	0	0	0	0	0
1473	0	1	0	0	0	0	0	0	0	0
1478	0	0	0	0	0	0	0	0	0	0
1480	0	0	0	0	0	0	0	0	0	0
1485	0	1	0	0	0	0	0	0	0	0
1486	0	1	0	0	0	0	0	0	0	0
1489	0	1	0	0	0	0	0	0	0	0
1492	0	1	0	0	0	0	0	0	0	0
1495	0	1	0	0	0	0	0	0	0	0
1502	0	1	0	0	0	0	0	0	0	0
1504	0	1	0	0	0	0	0	0	0	0
1507	0	0	0	0	0	0	0	0	0	0
1511	0	1	0	0	0	0	0	0	0	0
1512	0	1	0	0	0	0	0	0	0	0
1514	0	1	0	0	0	0	0	0	0	0
1523	0	1	0	0	0	0	0	0	0	0
1538	0	0	0	0	0	0	0	0	0	0
1539	0	0	0	0	0	0	0	0	0	0
1542	0	0	0	0	0	0	0	0	0	0
1545	0	1	0	0	0	0	0	0	0	0
1546	0	1	0	0	0	0	1	1	0	0
1549	0	1	0	0	0	0	0	0	0	0



In []:

保険薬局回答データの解析を行う。そのために、まず、保険薬局の回答データの読み込みを行う。

In [2]:

```
df <- read.csv("D:/questionair/170418保險藥局回答.csv")  
df
```

id	q1	q2	q3	q4	q5	q6	q7_11t	q7_12t	q7_21t	...	q301_5	q301_6	q301_7	c
4	2	500	2	4	2	3	741	243	210	...	0	0	0	1
5	2	1210	2	4	2	4	432	284	78	...	0	0	0	C
6	1	1400	-1	6	3	4	1000	600	400	...	0	0	0	C
7	1	830	2	6	2	3	-1	-1	-1	...	0	0	0	C
8	2	1300	2	6	2	4	691	225	195	...	0	0	0	C
9	2	850	2	6	1	4	350	200	150	...	0	0	0	C
10	2	2200	2	6	3	4	500	300	300	...	1	1	1	1
11	2	200	2	3	2	3	331	95	81	...	0	0	0	C
12	1	992	2	6	2	4	1141	387	291	...	0	0	0	C
13	1	-1	2	6	2	4	-1	-1	-1	...	0	0	0	C
14	2	3815	2	6	5	4	1000	400	600	...	1	1	1	1
15	2	630	2	6	1	2	1054	495	160	...	0	0	0	C
16	1	2000	2	6	2	1	400	50	-1	...	1	1	1	1
17	2	3000	2	6	3	4	1100	200	150	...	0	0	0	C
18	2	1733	2	6	2	4	320	160	89	...	0	0	0	C
19	2	617	2	6	1	4	686	233	129	...	0	0	0	C
20	1	70	2	6	3	4	-1	-1	-1	...	0	0	0	C
21	1	1610	2	3	2	1	798	358	421	...	0	0	0	C
22	1	2200	2	6	3	4	1100	800	200	...	0	0	1	C
23	1	600	2	6	1	3	600	180	100	...	0	0	0	C
24	2	2000	2	6	2	4	-1	-1	-1	...	0	0	0	C
25	2	729	2	6	1	3	300	100	50	...	0	0	0	C
26	1	1000	2	6	2	4	1021	732	220	...	1	1	1	1
27	1	900	2	3	2	3	521	118	88	...	0	0	0	C
28	2	1800	2	6	3	2	1200	600	100	...	0	0	0	C
29	2	2407	2	6	3	3	1200	500	300	...	0	0	0	C
30	2	1400	2	6	2	4	760	245	180	...	0	0	0	C
31	2	1100	2	6	2	1	1200	350	50	...	0	0	0	C
32	1	542	2	6	1	4	820	300	120	...	0	0	0	C
33	2	3600	2	6	3	3	-1	-1	-1	...	0	0	0	C
...
1089	2	1127	2	6	1	4	1122	607	127	...	0	0	0	C
1090	1	1800	2	6	3	4	1200	600	100	...	0	0	0	C
1091	1	2200	2	6	4	4	1140	420	260	...	0	0	0	C

id	q1	q2	q3	q4	q5	q6	q7_11t	q7_12t	q7_21t	...	q301_5	q301_6	q301_7	c
1092	2	7075	2	6	5	1	1137	268	310	...	0	0	0	C
1093	1	734	2	4	1	4	650	235	250	...	0	0	0	C
1094	2	723	2	4	2	3	640	185	160	...	0	0	0	C
1095	1	2591	1	4	4	4	1650	-1	460	...	0	0	0	1
1096	2	700	2	6	2	1	1000	400	200	...	0	0	0	C
1097	1	1617	2	4	3	1	1200	200	200	...	0	0	0	C
1098	2	961	2	6	2	4	760	232	190	...	1	1	0	C
1099	2	933	2	6	2	3	516	133	123	...	0	0	0	C
1100	2	7000	2	6	5	4	1000	200	30	...	0	0	0	1
1101	2	1500	2	4	3	2	1100	300	400	...	1	1	0	1
1102	2	670	2	6	1	4	667	227	212	...	0	0	0	C
1103	2	800	2	4	1	3	1100	700	90	...	0	0	0	C
1104	1	1000	2	6	2	4	500	200	30	...	1	1	1	1
1105	1	689	2	6	2	4	-1	-1	-1	...	0	0	0	C
1106	1	2800	2	6	3	3	2200	700	300	...	0	0	0	C
1107	2	1500	2	6	2	4	1000	750	200	...	0	0	0	C
1108	2	3219	2	6	2	4	1020	342	180	...	0	0	0	C
1109	1	4600	2	6	5	4	1500	300	230	...	0	0	0	C
1110	1	600	2	6	1	2	1300	400	300	...	0	0	0	C
1111	2	688	2	6	1	4	900	400	300	...	1	1	1	C
1112	2	2000	2	6	3	3	-1	-1	-1	...	1	1	1	1
1113	1	1200	2	6	3	4	1150	800	50	...	0	0	0	1
1114	2	2200	2	6	1	1	600	200	30	...	0	0	0	C
1115	2	1000	2	6	1	4	730	700	20	...	0	0	0	C
1116	1	2500	2	6	4	3	900	400	300	...	0	0	0	C
1117	2	1527	2	6	2	4	-1	-1	-1	...	1	1	1	1
1118	2	1900	2	6	2	3	-1	-1	-1	...	0	0	0	1



回答した薬局についてグルーピングを行うため、「I 薬局に関する項目」の回答に対してクラスタリングを適用する。なお、データフレームの各列の意味は以下の通りである。

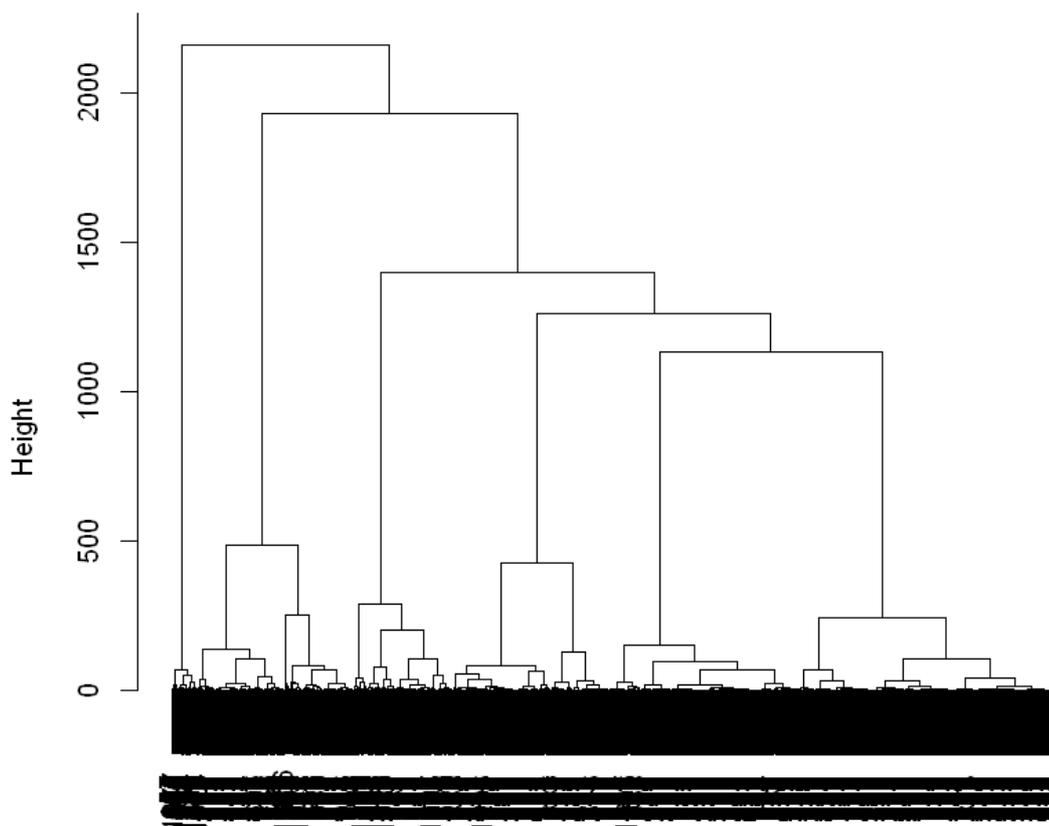
- q1:処方箋を応需する主たる医療機関の種別 (1:病院、2:診療所)
- q2:処方せん枚数(枚/月)
- q3:開設主体(1:個人、2:法人)
- q4:経営薬局数(1:1薬局、2:~10薬局、3:~20薬局、4:~50薬局、5:~99薬局、6:100~)
- q5:薬剤師数(常勤換算数)(1:1人、2:2~3人、3:4~6人、4:7~9人、5:10人~)
- q6:後発医薬品調剤割合(1:0~49%、2:50~64%、3:65~74%、4:75%~)

未回答は-1としている。q1~q6を未回答ありの回答を除外した上で階層的クラスタリングを適用する。

In [3]:

```
df1 <- df[df$q1>0 & df$q2>0 & df$q3>0 & df$q4>0 & df$q5>0,]  
z<-df1[,2:7]  
za<-scale(z)  
id<-df1[,1]  
z.d <- dist(za)^2  
result_cluster <- hclust(z.d, method="ward.D")  
plot(result_cluster)
```

Cluster Dendrogram



z.d
hclust (*, "ward.D")

得られたグラフ（デンドログラム）は大きく6つのクラスタに分けられうるとみなすことができる。そのため、このデンドログラムからクラスタを $k=6$ 個に分割する。

In [4]:

```
c1<-cutree(result_cluster, k=6)
(z1<-cbind(z, c1))
z2<-cbind(id, z1)
z_cluster <- z2[, c(1, 8)]
```

	q1	q2	q3	q4	q5	q6	c1
1	2	500	2	4	2	3	1
2	2	1210	2	4	2	4	1
4	1	830	2	6	2	3	2
5	2	1300	2	6	2	4	3
6	2	850	2	6	1	4	3
7	2	2200	2	6	3	4	4
8	2	200	2	3	2	3	1
9	1	992	2	6	2	4	2
11	2	3815	2	6	5	4	4
12	2	630	2	6	1	2	5
13	1	2000	2	6	2	1	5
14	2	3000	2	6	3	4	4
15	2	1733	2	6	2	4	3
16	2	617	2	6	1	4	3
17	1	70	2	6	3	4	2
18	1	1610	2	3	2	1	1
19	1	2200	2	6	3	4	2
20	1	600	2	6	1	3	2
21	2	2000	2	6	2	4	3
22	2	729	2	6	1	3	3
23	1	1000	2	6	2	4	2
24	1	900	2	3	2	3	1
25	2	1800	2	6	3	2	4
26	2	2407	2	6	3	3	4
27	2	1400	2	6	2	4	3
28	2	1100	2	6	2	1	5
29	1	542	2	6	1	4	2
30	2	3600	2	6	3	3	4
31	1	1000	2	6	2	4	2
32	2	610	2	6	1	4	3
...
1071	2	1127	2	6	1	4	3
1072	1	1800	2	6	3	4	2
1073	1	2200	2	6	4	4	4

	q1	q2	q3	q4	q5	q6	c1
1074	2	7075	2	6	5	1	4
1075	1	734	2	4	1	4	1
1076	2	723	2	4	2	3	1
1077	1	2591	1	4	4	4	6
1078	2	700	2	6	2	1	5
1079	1	1617	2	4	3	1	1
1080	2	961	2	6	2	4	3
1081	2	933	2	6	2	3	3
1082	2	7000	2	6	5	4	4
1083	2	1500	2	4	3	2	1
1084	2	670	2	6	1	4	3
1085	2	800	2	4	1	3	1
1086	1	1000	2	6	2	4	2
1087	1	689	2	6	2	4	2
1088	1	2800	2	6	3	3	4
1089	2	1500	2	6	2	4	3
1090	2	3219	2	6	2	4	3
1091	1	4600	2	6	5	4	4
1092	1	600	2	6	1	2	5
1093	2	688	2	6	1	4	3
1094	2	2000	2	6	3	3	4
1095	1	1200	2	6	3	4	2
1096	2	2200	2	6	1	1	5
1097	2	1000	2	6	1	4	3
1098	1	2500	2	6	4	3	4
1099	2	1527	2	6	2	4	3
1100	2	1900	2	6	2	3	3

クラスタの特徴を探るため、各クラスタにおける変数の平均値・分散について調べる。(結果のgroup:xは、クラスタxに相当) 結果より、それぞれのクラスタの特徴を鑑みて以下のように解釈することができる。

- クラスタ1:【中小規模】処方箋枚数約1400枚、診療所応需、法人、薬局数21～50、薬剤師2～3人、後発品65～74%
- クラスタ2:【病院】処方箋枚数約1200枚、病院応需、法人、薬局数100～、薬剤師2～3人、後発品75%～
- クラスタ3:【後発品多】処方箋枚数約1200枚、診療所応需、法人、薬局数100～、薬剤師2～3人、後発品75%～
- クラスタ4:【大規模】処方箋枚数約3300枚、診療所応需、法人、薬局数100～、薬剤師4～6人、後発品65～74%
- クラスタ5:【後発品少】処方箋枚数約1400枚、診療所応需、法人、薬局数100～、薬剤師2～3人、後発品50～64%
- クラスタ6:【個人】処方箋枚数約1600枚、診療所応需、個人、薬局数51～99、薬剤師2～3人、後発品65～74%

In [5]:

```
#options(repos="http://cran.ism.ac.jp/")  
#install.packages("psych", dependencies=TRUE)  
library(psych)  
describeBy(z1, group=c1)
```

Descriptive statistics by group

group: 1

	vars	n	mean	sd	median	trimmed	mad	min	max	range	skew	kurtosis
q1	1	120	1.65	0.48	2.0	1.69	0.00	1	2	1	-0.62	-1.63
q2	2	120	1464.73	759.54	1358.5	1434.99	809.50	35	3227	3192	0.34	-0.63
q3	3	120	2.00	0.00	2.0	2.00	0.00	2	2	0	NaN	NaN
q4	4	120	4.28	0.90	4.5	4.44	0.74	1	5	4	-1.40	1.95
q5	5	120	2.17	0.61	2.0	2.21	0.00	1	3	2	-0.10	-0.49
q6	6	120	3.18	0.85	3.0	3.31	1.48	1	4	3	-1.00	0.56
c1	7	120	1.00	0.00	1.0	1.00	0.00	1	1	0	NaN	NaN

se

q1	0.04
q2	69.34
q3	0.00
q4	0.08
q5	0.06
q6	0.08
c1	0.00

group: 2

	vars	n	mean	sd	median	trimmed	mad	min	max	range	skew	kurtosis
q1	1	229	1.00	0.00	1	1.00	0.00	1	1	0	NaN	NaN
q2	2	229	1189.84	582.24	1200	1187.77	557.46	32	2800	2768	0.13	-0.23
q3	3	229	2.00	0.00	2	2.00	0.00	2	2	0	NaN	NaN
q4	4	229	6.00	0.00	6	6.00	0.00	6	6	0	NaN	NaN
q5	5	229	2.02	0.61	2	2.01	0.00	1	4	3	0.34	0.82
q6	6	229	3.66	0.51	4	3.71	0.00	2	4	2	-1.03	-0.12
c1	7	229	2.00	0.00	2	2.00	0.00	2	2	0	NaN	NaN

se

q1	0.00
q2	38.48
q3	0.00
q4	0.00
q5	0.04
q6	0.03
c1	0.00

group: 3

	vars	n	mean	sd	median	trimmed	mad	min	max	range	skew
q1	1	311	2.00	0.00	2	2.00	0.00	2.0	2	0.0	NaN
q2	2	311	1191.78	626.51	1117	1158.07	524.84	33.3	3972	3938.7	0.72
q3	3	311	2.00	0.00	2	2.00	0.00	2.0	2	0.0	NaN
q4	4	311	5.94	0.23	6	6.00	0.00	5.0	6	1.0	-3.77
q5	5	311	1.71	0.47	2	1.75	0.00	1.0	3	2.0	-0.72
q6	6	311	3.61	0.49	4	3.64	0.00	3.0	4	1.0	-0.45
c1	7	311	3.00	0.00	3	3.00	0.00	3.0	3	0.0	NaN

kurtosis

se

q1	NaN	0.00
q2	1.28	35.53
q3	NaN	0.00
q4	12.24	0.01
q5	-0.99	0.03
q6	-1.80	0.03
c1	NaN	0.00

group: 4

	vars	n	mean	sd	median	trimmed	mad	min	max	range	skew
q1	1	187	1.57	0.50	2	1.58	0.00	1	2	1	-0.27
q2	2	187	3269.97	1492.68	3000	3085.09	1005.20	900	11000	10100	1.89
q3	3	187	2.00	0.00	2	2.00	0.00	2	2	0	NaN
q4	4	187	6.00	0.00	6	6.00	0.00	6	6	0	NaN

```

q5  5 187  3.41  0.72  3  3.29  0.00  2  5  3  1.09
q6  6 187  3.09  0.88  3  3.21  1.48  1  4  3 -0.84
c1  7 187  4.00  0.00  4  4.00  0.00  4  4  0  NaN
      kurtosis      se
q1  -1.94  0.04
q2   5.57 109.16
q3   NaN  0.00
q4   NaN  0.00
q5   0.24  0.05
q6   0.10  0.06
c1   NaN  0.00

```

group: 5

```

      vars  n   mean    sd median trimmed  mad min  max range  skew kurtosis
q1     1 187   1.62  0.49     2   1.65   0.00  1   2     1 -0.49   -1.77
q2     2 187 1406.89 810.96  1281 1333.78 680.51 38 4163 4125  1.02    1.42
q3     3 187   2.00  0.00     2   2.00   0.00  2   2     0  NaN     NaN
q4     4 187   5.91  0.30     6   6.00   0.00  4   6     2 -3.52   12.54
q5     5 187   1.90  0.64     2   1.87   0.00  1   4     3  0.33    0.31
q6     6 187   1.47  0.62     2   1.52   0.00 -1   2     3 -1.26    2.97
c1     7 187   5.00  0.00     5   5.00   0.00  5   5     0  NaN     NaN
      se
q1  0.04
q2 59.30
q3  0.00
q4  0.02
q5  0.05
q6  0.05
c1  0.00

```

group: 6

```

      vars  n   mean    sd median trimmed  mad min  max range  skew kurtosis
q1     1  31   1.74  0.44     2   1.80   0.00  1   2     1 -1.05   -0.92
q2     2  31 1626.81 776.14  1800 1643.24 741.30 80 3000 2920 -0.31   -1.02
q3     3  31   1.00  0.00     1   1.00   0.00  1   1     0  NaN     NaN
q4     4  31   5.35  0.71     5   5.44   1.48  4   6     2 -0.58   -0.93
q5     5  31   2.06  0.81     2   2.00   0.00  1   4     3  0.61    0.01
q6     6  31   3.16  0.93     3   3.28   1.48  1   4     3 -0.78   -0.48
c1     7  31   6.00  0.00     6   6.00   0.00  6   6     0  NaN     NaN
      se
q1  0.08
q2 139.40
q3  0.00
q4  0.13
q5  0.15
q6  0.17
c1  0.00

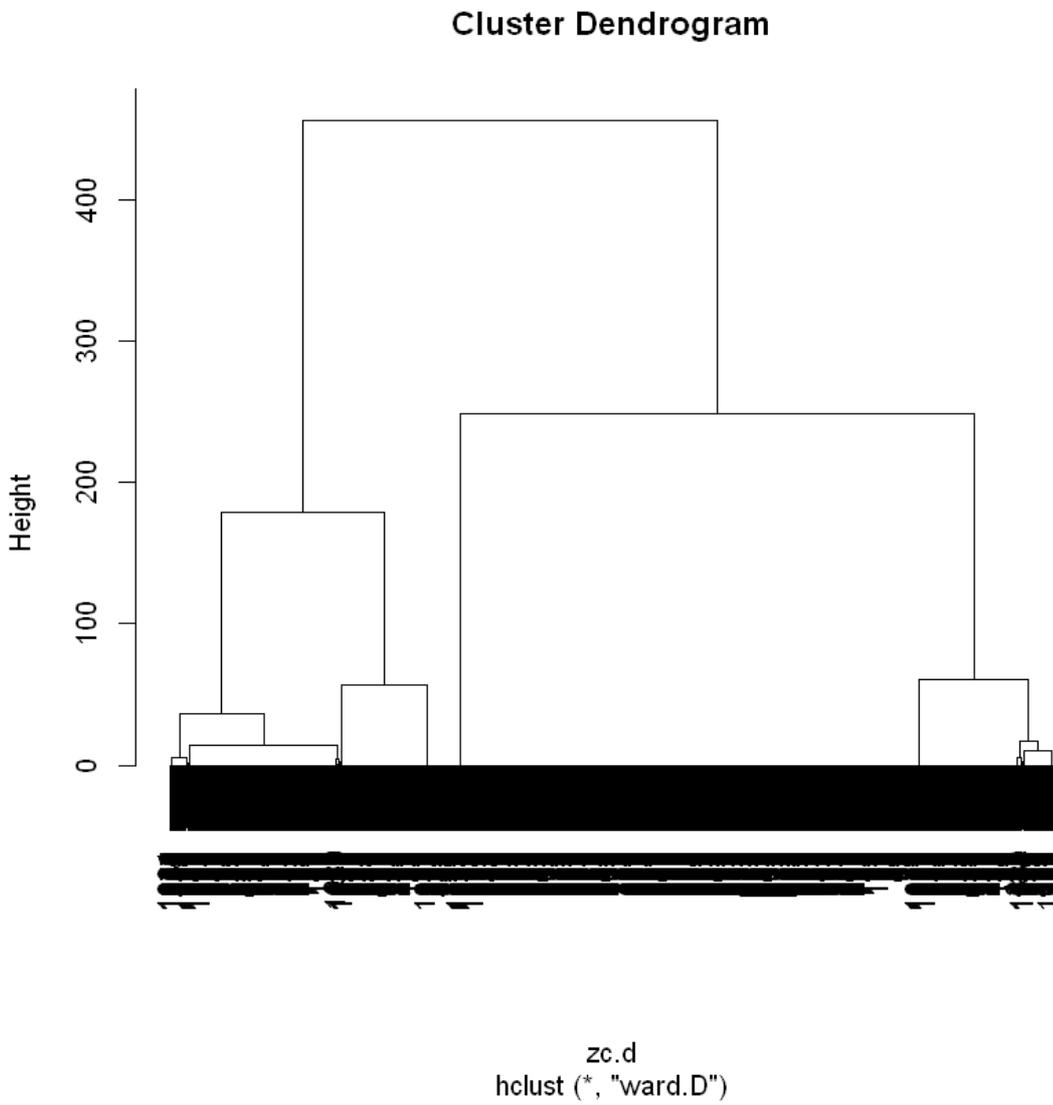
```

In [6]:

```
clusteredData<-merge(z_cluster, df, by="id")
```

In [7]:

```
zc<-clusteredData[,c("q105_1","q105_2","q105_3","q105_4","q105_5","q105_6")]  
idc <- clusteredData[,1:2]  
zc.d <- dist(zc)^2  
result_cluster_q105 <- hclust(zc.d, method="ward.D")  
plot(result_cluster_q105)
```



In [8]:

```
c_q105<-cutree(result_cluster_q105, k=4)
(z1c<-cbind(zc, c_q105))
```

q105_1	q105_2	q105_3	q105_4	q105_5	q105_6	c_q105
0	0	0	0	0	0	1
1	1	0	0	0	0	2
1	1	0	0	0	0	2
1	1	0	0	0	0	2
1	1	0	0	0	0	2
1	1	1	0	0	0	3
0	0	0	0	0	0	1
1	1	0	0	0	0	2
1	1	0	0	0	0	2
1	1	0	0	0	0	2
0	1	0	0	0	0	4
1	1	1	0	0	0	3
1	1	0	0	0	0	2
1	1	0	0	0	0	2
1	1	0	0	0	0	2
0	0	0	0	0	0	1
1	1	0	0	0	0	2
1	1	0	0	0	0	2
0	0	0	0	0	0	1
1	1	1	0	0	0	3
1	1	0	0	0	0	2
0	1	0	0	0	0	4
0	0	0	0	0	1	4
1	1	0	0	0	0	2
1	1	1	0	0	0	3
1	1	0	0	0	0	2
0	1	1	0	0	0	4
1	1	1	0	0	0	3
0	1	0	0	0	0	4
0	1	0	0	0	0	4
...
1	1	0	0	0	0	2
1	1	0	0	0	0	2
0	1	0	0	0	0	4

q105_1	q105_2	q105_3	q105_4	q105_5	q105_6	c_q105
0	1	0	0	0	0	4
0	1	0	0	0	0	4
0	0	0	0	0	0	1
1	1	1	1	0	0	3
0	0	0	0	0	0	1
1	1	1	1	0	0	3
1	0	0	0	0	0	1
0	1	0	0	0	0	4
1	1	0	0	0	0	2
0	1	0	0	0	0	4
0	1	0	0	0	0	4
0	1	0	0	0	0	4
1	1	0	1	0	0	3
1	1	0	0	0	0	2
1	1	0	0	0	0	2
1	1	0	0	0	0	2
0	1	0	0	0	0	4
1	1	1	0	0	1	4
0	1	0	0	0	0	4
1	1	0	0	0	0	2
1	1	1	0	0	0	3
1	1	1	0	0	0	3
1	1	1	0	0	0	3
1	1	0	0	0	0	2
0	1	0	0	0	0	4
1	1	0	0	0	0	2
1	1	0	0	0	0	2

薬剤部門にある医薬品に関するバーコードリーダー有無についてのクラスタの特徴を探るため、各クラスタにおける変数の平均値・分散について調べる。(結果のgroup:xは、クラスタxに相当) 結果より、それぞれのクラスタの特徴を鑑みて以下のように解釈することができる。

- クラスタ1:チェックなし
- クラスタ2:薬剤管理部門と調剤部門
- クラスタ3:薬剤管理部門と調剤部門と注射薬調剤部門
- クラスタ4:薬剤管理部門のみ

In [10]:

```
describeBy(z1c, group=c_q105)
```

Descriptive statistics by group

```
group: 1
```

	vars	n	mean	sd	median	trimmed	mad	min	max	range	skew	kurtosis	se
q105_1	1	142	0.27	0.45	0	0.22	0	0	1	1	1	-1.01	0.04
q105_2	2	142	0.00	0.00	0	0.00	0	0	0	0	NaN	NaN	0.00
q105_3	3	142	0.00	0.00	0	0.00	0	0	0	0	NaN	NaN	0.00
q105_4	4	142	0.00	0.00	0	0.00	0	0	0	0	NaN	NaN	0.00
q105_5	5	142	0.00	0.00	0	0.00	0	0	0	0	NaN	NaN	0.00
q105_6	6	142	0.00	0.00	0	0.00	0	0	0	0	NaN	NaN	0.00
c_q105	7	142	1.00	0.00	1	1.00	0	1	1	0	NaN	NaN	0.00

```
group: 2
```

	vars	n	mean	sd	median	trimmed	mad	min	max	range	skew	kurtosis	se
q105_1	1	552	1	0	1	1	0	1	1	0	NaN	NaN	0
q105_2	2	552	1	0	1	1	0	1	1	0	NaN	NaN	0
q105_3	3	552	0	0	0	0	0	0	0	0	NaN	NaN	0
q105_4	4	552	0	0	0	0	0	0	0	0	NaN	NaN	0
q105_5	5	552	0	0	0	0	0	0	0	0	NaN	NaN	0
q105_6	6	552	0	0	0	0	0	0	0	0	NaN	NaN	0
c_q105	7	552	2	0	2	2	0	2	2	0	NaN	NaN	0

```
group: 3
```

	vars	n	mean	sd	median	trimmed	mad	min	max	range	skew	kurtosis	se
q105_1	1	166	0.99	0.08	1	1.0	0	0	1	1	-12.65	159.04	0.01
q105_2	2	166	1.00	0.00	1	1.0	0	1	1	0	NaN	NaN	0.00
q105_3	3	166	0.95	0.21	1	1.0	0	0	1	1	-4.18	15.57	0.02
q105_4	4	166	0.26	0.44	0	0.2	0	0	1	1	1.09	-0.82	0.03
q105_5	5	166	0.05	0.21	0	0.0	0	0	1	1	4.18	15.57	0.02
q105_6	6	166	0.00	0.00	0	0.0	0	0	0	0	NaN	NaN	0.00
c_q105	7	166	3.00	0.00	3	3.0	0	3	3	0	NaN	NaN	0.00

```
group: 4
```

	vars	n	mean	sd	median	trimmed	mad	min	max	range	skew	kurtosis	se
q105_1	1	205	0.01	0.10	0	0.00	0	0	1	1	9.90	96.53	0.01
q105_2	2	205	0.99	0.10	1	1.00	0	0	1	1	-9.90	96.53	0.01
q105_3	3	205	0.11	0.31	0	0.01	0	0	1	1	2.52	4.37	0.02
q105_4	4	205	0.01	0.12	0	0.00	0	0	1	1	8.02	62.70	0.01
q105_5	5	205	0.00	0.00	0	0.00	0	0	0	0	NaN	NaN	0.00
q105_6	6	205	0.03	0.17	0	0.00	0	0	1	1	5.54	28.88	0.01
c_q105	7	205	4.00	0.00	4	4.00	0	4	4	0	NaN	NaN	0.00

In [11]:

```
z2c<-cbind(idc, z1c)
z3c<-z2c[, c("id", "c_q105")]
clusteredData2<-merge(z3c, clusteredData, by="id")
z_final<-clusteredData2[clusteredData2$q106_1>0, ]
#z_final$q106_1[z_final$q106_1==2]<-1
```

In [12]:

```
z_final_for_cross=z_final[,c("c1", "c_q105")]
xtabs(~c1+c_q105, z_final_for_cross)
```

```
  c_q105
c1   1   2   3   4
  1  47   9   7  43
  2  11 127  39  29
  3  26 159  36  52
  4  18  96  31  28
  5  17 104  24  23
  6   9   6   4   8
```

In [14]:

```
#options(repos="http://cran.ism.ac.jp/")
#install.packages("partykit", dependencies=TRUE)
#install.packages("rpart.plot", dependencies=TRUE)
```

In [14]:

```
library(rpart)
library(rpart.plot)
library(partykit)
```

決定木による解析では、「計画はない」とする回答が多く、明確な結果が得られなかった。

In [16]:

```
ct <- rpart(q106_1~q1+q2+q3+q4+q5+q6+q101+c_q105, data = z_final, method='class')
print(ct)
```

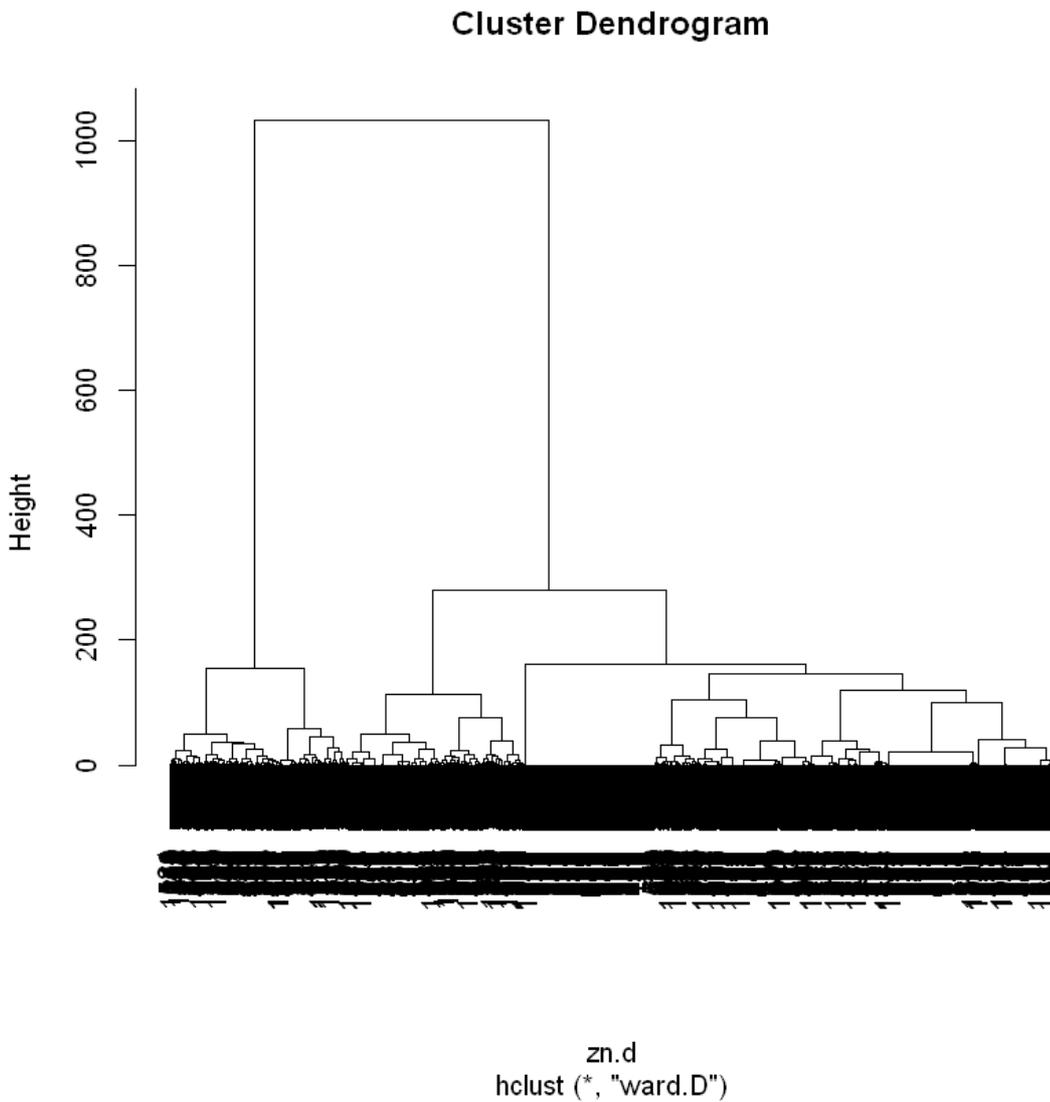
n= 953

```
node), split, n, loss, yval, (yprob)
  * denotes terminal node
```

```
1) root 953 192 3 (0.05351522 0.14795383 0.79853095) *
```

In [17]:

```
z_needs<-z_final[, c("q301_1", "q301_2", "q301_3", "q301_4", "q301_5", "q301_6", "q301_7", "q301_8", "q301_9", "q301_10", "q301_11", "q301_12", "q301_13", "q301_14")]  
id<-z_final[, 1]  
zn.d <- dist(z_needs)^2  
result_needs_cluster <- hclust(zn.d, method="ward.D")  
plot(result_needs_cluster)
```



In [18]:

```
c_needs<-cutree(result_needs_cluster, k=3)
(z1needs<-cbind(z_needs, c_needs))
```

	q301_1	q301_2	q301_3	q301_4	q301_5	q301_6	q301_7	q301_8	q301_9	q3
1	1	1	0	0	0	0	0	1	1	0
2	0	1	0	0	0	0	0	0	0	0
3	0	0	0	0	0	0	0	0	0	0
4	1	1	0	0	0	0	0	0	0	0
5	1	1	0	0	0	0	0	0	0	0
6	1	1	1	1	1	1	1	1	1	1
7	1	0	1	0	0	0	0	0	0	0
8	0	1	0	0	0	0	0	0	0	0
9	0	1	1	1	1	1	1	1	1	0
12	0	1	0	0	0	0	0	0	0	0
13	0	1	0	0	0	0	0	0	0	0
16	0	0	0	0	0	0	0	0	0	0
17	0	1	0	0	0	0	1	0	0	0
18	0	0	1	0	0	0	0	0	0	0
20	1	1	0	0	0	0	0	0	0	0
21	0	1	0	1	1	1	1	1	1	0
22	0	0	0	0	0	0	0	0	0	0
23	1	1	1	0	0	0	0	0	0	0
24	0	0	0	0	0	0	0	0	0	0
27	0	1	0	0	0	0	0	0	0	0
28	0	0	0	0	0	0	0	0	0	0
29	0	0	0	0	0	1	0	0	0	0
30	0	0	0	0	0	0	0	0	0	0
31	1	1	0	0	0	0	0	0	0	0
32	1	1	1	1	1	1	1	1	1	0
33	1	1	1	1	0	0	0	0	0	0
34	0	1	0	0	0	0	0	0	0	0
35	0	1	0	0	0	0	0	0	0	0
37	1	0	0	0	0	0	0	0	0	0
38	0	1	0	1	1	1	1	0	0	0
...
1033	1	0	1	0	0	0	1	0	0	0
1034	0	1	0	0	0	0	0	0	0	0
1035	0	0	0	0	0	0	0	0	0	0

	q301_1	q301_2	q301_3	q301_4	q301_5	q301_6	q301_7	q301_8	q301_9	q3
1036	0	1	1	0	0	0	0	0	0	0
1037	0	0	0	0	0	0	0	0	0	0
1038	0	1	0	0	0	0	0	0	0	0
1039	0	1	1	0	0	0	0	0	0	0
1040	0	1	1	0	0	0	0	0	0	0
1041	0	0	0	0	0	0	0	0	0	0
1042	0	1	0	0	0	0	0	1	0	0
1043	0	1	0	0	0	0	0	0	0	0
1044	1	0	0	0	0	0	0	0	0	0
1045	1	1	1	1	1	1	0	0	0	0
1046	0	1	0	0	0	0	0	0	0	0
1047	0	0	1	0	0	0	0	1	1	1
1048	0	0	1	1	1	1	0	1	0	0
1049	0	1	0	0	0	0	0	0	0	0
1050	0	1	1	0	0	0	0	0	0	0
1051	1	1	1	1	1	1	1	1	1	1
1053	0	0	0	0	0	0	0	0	0	0
1054	1	1	0	0	0	0	0	0	1	0
1055	0	0	0	0	0	0	0	0	0	0
1056	1	0	0	0	0	0	0	0	0	0
1057	0	0	0	0	0	0	0	0	0	0
1058	1	1	0	1	1	1	1	0	0	1
1059	1	1	0	1	1	1	1	1	0	1
1062	0	0	0	0	0	0	0	0	0	0
1063	0	0	1	0	0	0	0	0	0	0
1064	1	1	1	1	1	1	1	1	1	1
1065	0	1	0	0	0	0	0	1	1	0

バーコードを変更してほしい事例についての3つのクラスタについては以下のように解釈できる。

- クラスタ1:回答なし
- クラスタ2:ブランド名の変更、製剤の色・形・大きさ・識別記号の変更、管理保管方法変更時
- クラスタ3:ブランド名の変更、管理保管方法変更時、銘柄別⇔統一名収載変更時

In [19]:

```
describeBy(z_needs, group=c_needs)
```

Descriptive statistics by group

```
group: 1
```

	vars	n	mean	sd	median	trimmed	mad	min	max	range	skew	kurtosis	se
q301_1	1	572	0.25	0.43	0	0.19	0	0	1	1	1.15	-0.67	0.02
q301_2	2	572	0.50	0.50	0	0.50	0	0	1	1	0.01	-2.00	0.02
q301_3	3	572	0.15	0.36	0	0.06	0	0	1	1	1.95	1.81	0.01
q301_4	4	572	0.01	0.12	0	0.00	0	0	1	1	8.26	66.27	0.00
q301_5	5	572	0.00	0.04	0	0.00	0	0	1	1	23.79	565.01	0.00
q301_6	6	572	0.01	0.08	0	0.00	0	0	1	1	11.80	137.51	0.00
q301_7	7	572	0.02	0.13	0	0.00	0	0	1	1	7.34	52.02	0.01
q301_8	8	572	0.08	0.26	0	0.00	0	0	1	1	3.21	8.34	0.01
q301_9	9	572	0.06	0.24	0	0.00	0	0	1	1	3.65	11.36	0.01
q301_10	10	572	0.02	0.15	0	0.00	0	0	1	1	6.39	38.88	0.01
q301_11	11	572	0.28	0.45	0	0.22	0	0	1	1	1.00	-1.01	0.02
q301_12	12	572	0.12	0.33	0	0.03	0	0	1	1	2.27	3.18	0.01
q301_13	13	572	0.02	0.14	0	0.00	0	0	1	1	6.67	42.53	0.01
q301_14	14	572	0.02	0.14	0	0.00	0	0	1	1	6.98	46.84	0.01

```
group: 2
```

	vars	n	mean	sd	median	trimmed	mad	min	max	range	skew	kurtosis	se
q301_1	1	188	0.46	0.50	0	0.45	0	0	1	1	0.15	-1.99	0.04
q301_2	2	188	0.77	0.42	1	0.83	0	0	1	1	-1.25	-0.45	0.03
q301_3	3	188	0.49	0.50	0	0.49	0	0	1	1	0.04	-2.01	0.04
q301_4	4	188	0.95	0.23	1	1.00	0	0	1	1	-3.95	13.68	0.02
q301_5	5	188	0.98	0.13	1	1.00	0	0	1	1	-7.66	57.04	0.01
q301_6	6	188	0.88	0.32	1	0.97	0	0	1	1	-2.36	3.61	0.02
q301_7	7	188	0.66	0.47	1	0.70	0	0	1	1	-0.69	-1.53	0.03
q301_8	8	188	0.36	0.48	0	0.33	0	0	1	1	0.57	-1.68	0.04
q301_9	9	188	0.24	0.43	0	0.18	0	0	1	1	1.18	-0.61	0.03
q301_10	10	188	0.21	0.41	0	0.14	0	0	1	1	1.39	-0.06	0.03
q301_11	11	188	0.49	0.50	0	0.49	0	0	1	1	0.04	-2.01	0.04
q301_12	12	188	0.54	0.50	1	0.55	0	0	1	1	-0.15	-1.99	0.04
q301_13	13	188	0.24	0.43	0	0.18	0	0	1	1	1.18	-0.61	0.03
q301_14	14	188	0.37	0.48	0	0.34	0	0	1	1	0.52	-1.73	0.04

```
group: 3
```

	vars	n	mean	sd	median	trimmed	mad	min	max	range	skew	kurtosis	se
q301_1	1	193	0.29	0.46	0	0.24	0	0	1	1	0.92	-1.16	0.03
q301_2	2	193	0.67	0.47	1	0.71	0	0	1	1	-0.71	-1.50	0.03
q301_3	3	193	0.33	0.47	0	0.28	0	0	1	1	0.73	-1.47	0.03
q301_4	4	193	0.05	0.22	0	0.00	0	0	1	1	4.01	14.18	0.02
q301_5	5	193	0.08	0.28	0	0.00	0	0	1	1	3.00	7.05	0.02
q301_6	6	193	0.01	0.07	0	0.00	0	0	1	1	13.68	186.03	0.01
q301_7	7	193	0.13	0.34	0	0.04	0	0	1	1	2.19	2.81	0.02
q301_8	8	193	0.11	0.31	0	0.01	0	0	1	1	2.49	4.24	0.02
q301_9	9	193	0.05	0.21	0	0.00	0	0	1	1	4.27	16.29	0.02
q301_10	10	193	0.10	0.30	0	0.00	0	0	1	1	2.67	5.18	0.02
q301_11	11	193	0.49	0.50	0	0.48	0	0	1	1	0.05	-2.01	0.04
q301_12	12	193	0.52	0.50	1	0.53	0	0	1	1	-0.09	-2.00	0.04
q301_13	13	193	0.28	0.45	0	0.23	0	0	1	1	0.97	-1.06	0.03
q301_14	14	193	0.80	0.40	1	0.87	0	0	1	1	-1.47	0.17	0.03

In [20]:

```
z_needs<-cbind(z_final,c_needs)
z_needs_final<-z_needs[z_needs$q101>0 & z_needs$q106_1>0 & z_needs$q1>0 & z_needs$q2>0 &
z_needs$q3>0 & z_needs$q4>0 & z_needs$q5>0 & z_needs$q6>0
& z_needs$c_q105>0,]
```

In [21]:

```
ct3 <- rpart(c_needs~q1+q2+q3+q4+q5+q6+q101+c_q105, data = z_needs_final, method='class')
print(ct3)
```

n= 944

node), split, n, loss, yval, (yprob)
* denotes terminal node

1) root 944 377 1 (0.6006356 0.1970339 0.2023305) *

決定木分析の結果より、未チェックの回答が多く明確な結果が得られなかった。

In [22]:

```
z_final_needs_for_cross=z_needs[,c("c_needs", "c1")]
xtabs(~c1+c_needs,z_final_needs_for_cross)
```

	c_needs		
c1	1	2	3
1	64	22	20
2	129	42	35
3	168	51	54
4	105	27	41
5	90	42	36
6	16	4	7

