

データマイニング入門

第6章 クラスター分析 —Rによるクラスター分析

10NM733X

林華

6.3.1 動物の分類—動物データ

- > library(MASS)
- > data(Animals)
- > 動物データ3 <- Animals[c(2, 9, 13, 14, 24),]
- > rownames(動物データ3) <- c("牛", "馬", "ゴリラ", "人間", "チンパンジー")
- > colnames(動物データ3) <- c("体重", "脳の重さ")
- > 動物データ3

	体重	脳の重さ
牛	465.00	423
馬	521.00	655
ゴリラ	207.00	406
人間	62.00	1320
チンパンジー	52.16	440

>

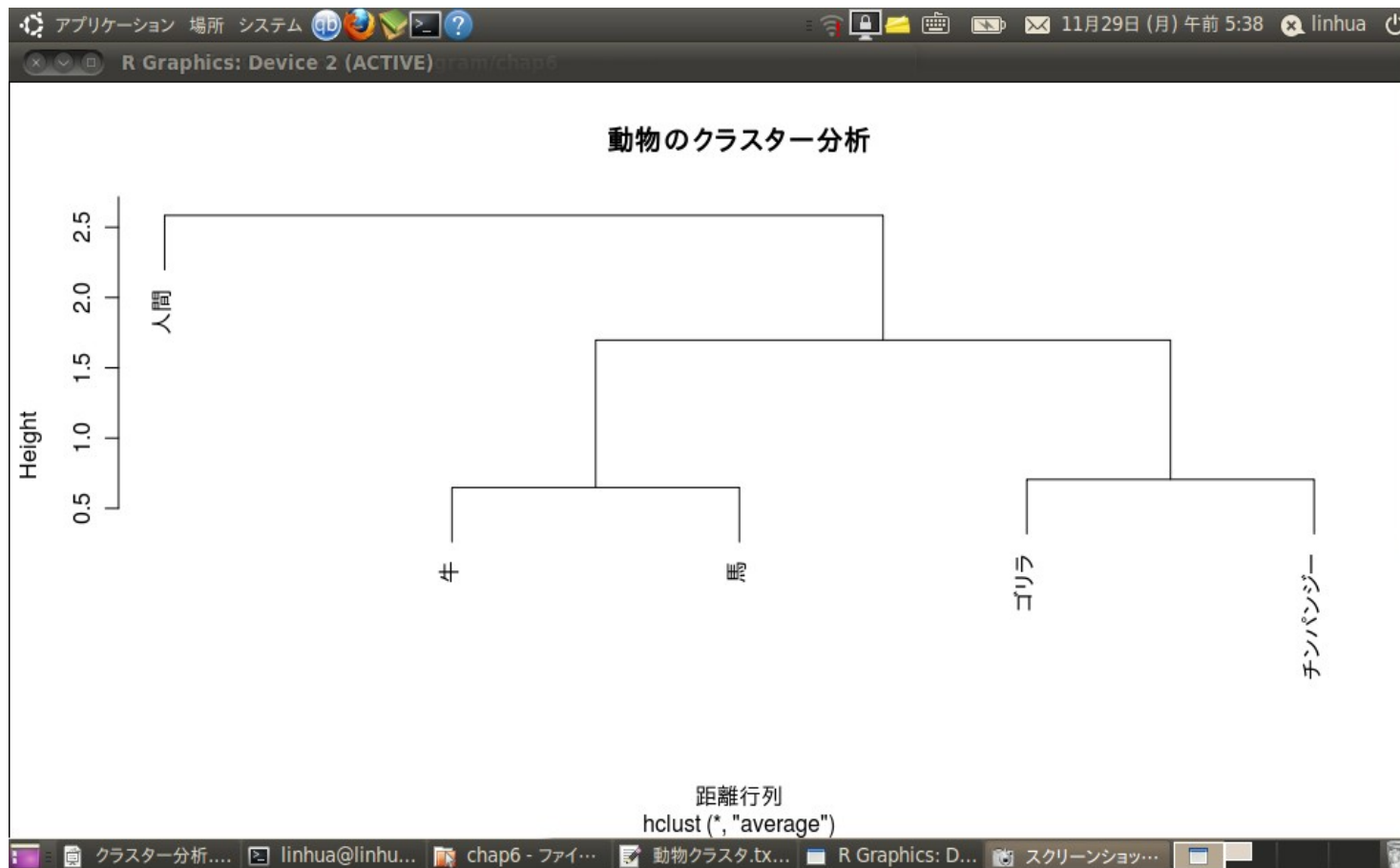
6.3.1 動物データの距離行列

```
> 距離行列 <- dist(scale(動物データ3))  
> round(距離行列, 3)
```

	牛	馬	ゴリラ	人間
馬	0.649			
ゴリラ	1.168	1.559		
人間	2.942	2.691	2.442	
チンパンジー	1.869	2.192	0.706	2.265

6.3.1 動物クラスター

- 動物クラスター ← `hclust(距離行列, method="average")`
- `plot(動物クラスター, hang=0.2, main="動物のクラスター分析")`



6.3.2 街のイメージ調査

> 街データ <- read.csv("街のイメージ.csv", header=T, row.names=1)

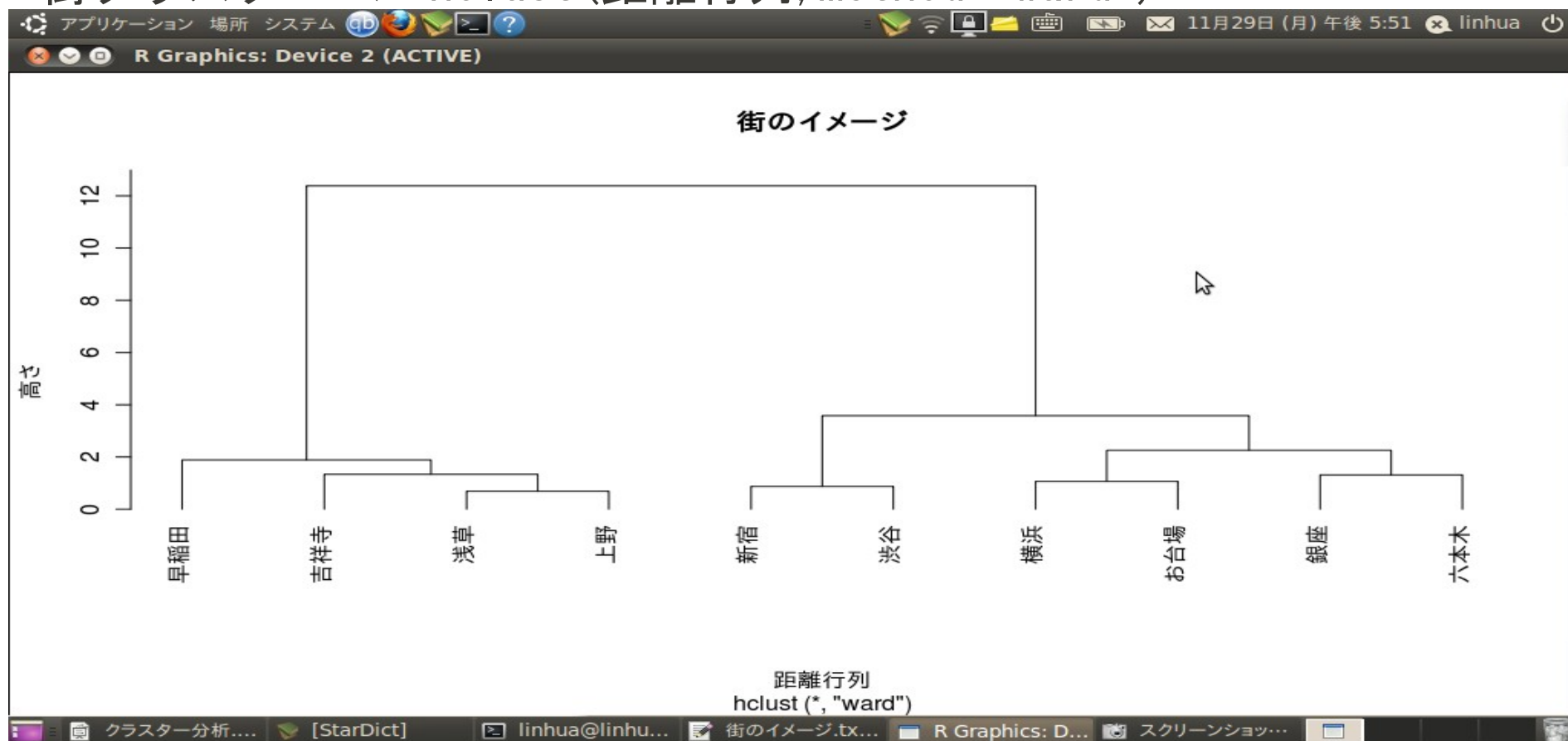
> 街データ

	派手	開放的	冷静	しゃれた	いそがしい	先進的
新宿	4.373	3.542	2.203	3.271	4.712	4.153
横浜	3.305	3.814	2.864	4.136	3.322	3.831
吉祥寺	2.085	3.153	3.017	3.441	2.153	2.492
銀座	4.017	2.475	3.458	4.576	3.322	3.814
浅草	1.932	2.932	2.898	2.458	2.322	1.746
渋谷	4.695	3.627	1.492	3.390	4.373	4.017
お台場	3.949	3.949	2.220	4.186	3.407	4.356
上野	2.068	3.169	3.136	2.508	2.610	2.254
六本木	4.407	3.034	2.932	4.559	4.000	4.542
早稲田	1.763	3.186	2.322	1.695	3.102	2.017

>

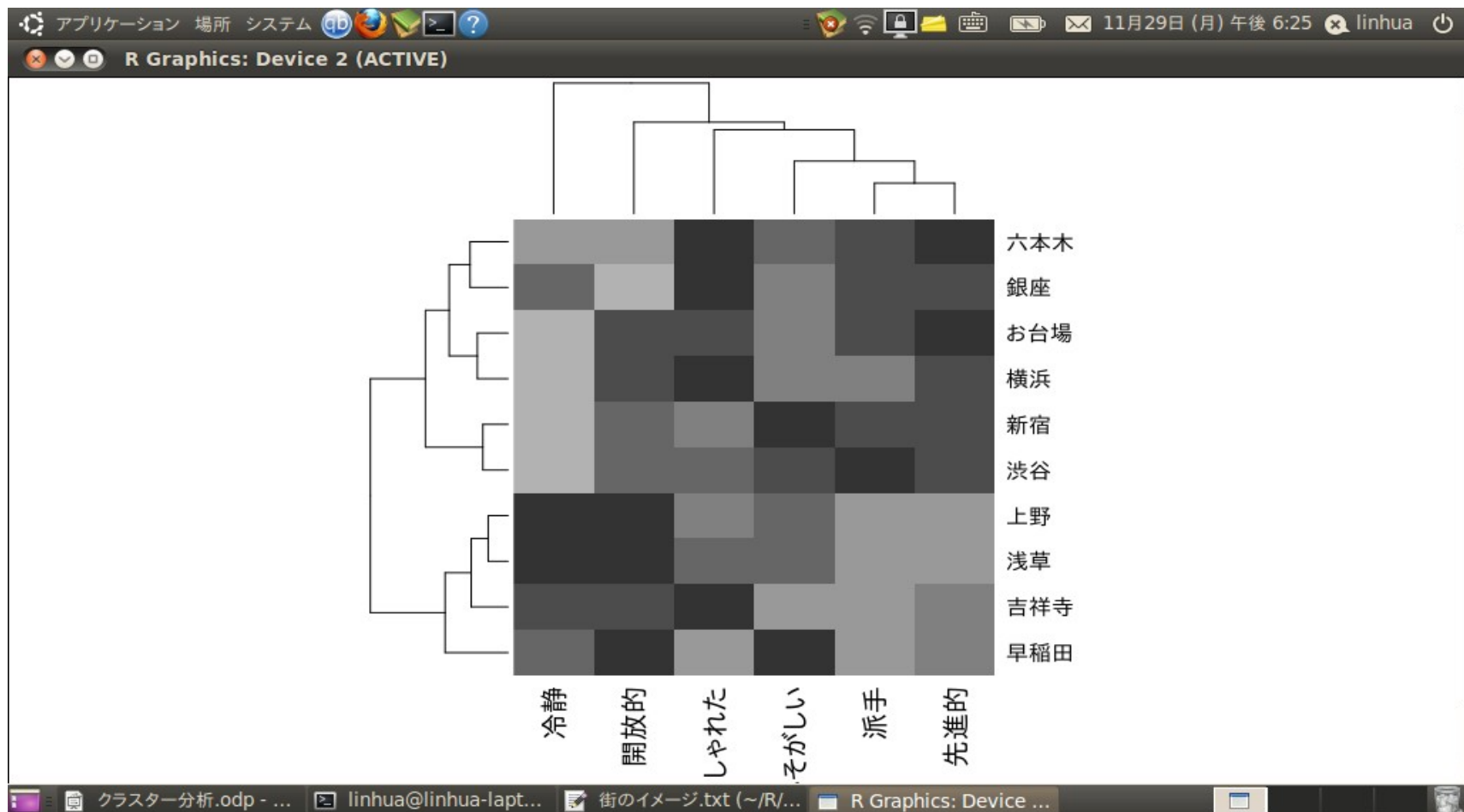
6.3.2 街のイメージーデンドログラム

- 街データ \leftarrow read.csv("街のイメージ.csv", header=T, row.names=1)
- 距離行列 \leftarrow dist(街データ, method="euclidean")
- 街クラスター \leftarrow hclust(距離行列, method="ward")



6.3.2 街のイメージ—ヒートマップ

- heatmap(as.matrix(街データ), col=c(gray(0.7), gray(0.6), gray(0.5), gray(0.4), gray(0.3), gray(0.2)))



6.3.3 居酒屋メニューの分類

> 居酒屋 <- read.csv("居酒屋.csv", header=T, row.names=1)

> 居酒屋

	こってり	遅さ	安さ	少なさ	馴染み
枝豆	1.680	1.448	4.142	2.808	4.840
ナスの漬物	1.429	1.662	4.013	3.481	4.117
きゅうり漬	1.348	1.478	4.043	3.551	4.246
漬物盛合せ	1.425	1.630	3.534	3.260	4.260
キムチ	2.705	1.769	3.795	3.410	4.154
冷奴	1.195	1.468	3.935	3.662	4.610
えびせん	3.319	1.971	3.826	2.754	3.507
揚げパスタ	3.726	2.945	3.000	2.781	2.466
もずく酢	1.295	1.872	3.641	3.705	2.910
鶏の唐揚	4.221	3.584	2.818	2.714	4.675

・
・
・

6.3.3 居酒屋メニューの分類—Kmeans

>居酒屋クラスター <-kmeans(居酒屋, centers=4)

>居酒屋クラスター

K-means clustering with 4 clusters of sizes 5, 8, 12, 8

Cluster means:

```
こってり 遅さ 安さ 少なさ 馴染み
1 1.752000 2.334400 2.645800 3.655200 3.25800
2 1.667875 1.718250 3.817750 3.282125 4.37300
3 3.557667 3.329917 2.912167 3.204417 3.50450
4 4.206625 3.366250 3.038250 2.425500 4.19675
```

Clustering vector:

枝豆	ナスの漬物	きゅうり漬け	漬物盛合せ	キムチ
2	2	2	2	2
冷奴	えびせん	揚げパスタ	もずく酢	鶏の唐揚
2	3	3	1	4
なんこつ(揚物)	チーズフライ	フライドポテト	焼き鳥	砂肝
4	3	4	3	3
ねぎま	つくね	アジの開き	たこぶつ刺身	マグロの刺身
3	3	3	1	1
イカの刺身	サーモン刺身	ほっけ焼き	一口ステーキ	ししゃも
1	1	3	3	3
チャーハン	焼きそば	焼きうどん	ピザ	お好み焼き
4	4	4	4	4
パスタ	お茶漬け	おにぎり		
3	2	2		

Within cluster sum of squares by cluster:

```
[1] 4.204263 4.069651 13.518009 4.876914
```

Available components:

```
[1] "cluster" "centers" "withinss" "size"
```

6.3.3 居酒屋メニューの分類—Kmeans

- > tapply(names(居酒屋クラスター\$cluster), 居酒屋クラスター\$cluster, unique)
- \$`1`
 - [1] “もずく酢” “たこぶつ刺身” “マグロの刺身” “イカの刺身” “サーモン刺身”
- \$`2`
 - [1] “枝豆” “ナスの漬物” “きゅうり漬け” “漬物盛合せ” “キムチ”
 - [6] “冷奴” “お茶漬け” “おにぎり”
- \$`3`
 - [1] “えびせん” “揚げパスタ” “チーズフライ” “焼き鳥” “砂肝”
 - [6] “ねぎま” “つくね” “アジの開き” “ほっけ焼き” “一口ステーキ”
 - [11] “ししゃも” “パスタ”
- \$`4`
 - [1] “鶏の唐揚” “なんこつ(揚物)” “フライドポテト” “チャーハン”
 - [5] “焼きそば” “焼きうどん” “ピザ” “お好み焼き”
- >

6.3.4 チロルチョコの再分析

```
> チロルデータ2 <- read.csv('チロル2.csv', header=T, row.names=1)
```

```
> チロルデータ2
```

	奇抜さ	リッチさ	くどさ	おしゃれ度	大人度	和風さ
コーヒーマスター	5	6	5	6	3	1
アーモンド	2	3	3	4	2	5
ストロベリー	3	3	3	2	2	3
ビス	4	2	4	3	1	2
きなこもち	5	3	4	3	5	6
冬チョコ	3	5	3	5	4	2
いちごびす	3	4	5	2	1	2
エッグタルト	4	5	6	5	4	3
パイ	3	4	2	4	2	3
バナナ	2	5	5	4	2	3
すいとんぽてと	5	6	6	5	4	3
抹茶	2	4	4	5	5	6
カプチーノ	4	5	3	5	6	2
黒ごまと	5	5	6	5	6	5
いちごプリン	2	3	5	3	2	2
みるく大福	5	4	6	4	5	5
.						
.						

6.3.4 チロルチョコの再分析

```
>チロルクラスタ<-kmeans(チロルデータ2, centers=5)
> tapply(names(チロルクラスタ$cluster), チロルクラスタ$cluster, unique)
$`1`
[1] "バナナ"          "キャラメル"      "ふわふわココア" "シナモンオレ"

$`2`
[1] "コーヒーヌガー" "エッグタルト"   "すいとぽてと"   "黒ごまとると"
[5] "みるく大福"     "杏仁豆腐"       "パンプキンタルト" "クレメダンジュ"

$`3`
[1] "きなこもち"     "抹茶"           "RICE"           "ゆず"
[5] "塩バニラアソート"

$`4`
[1] "アーモンド"   "ストロベリー" "ビス"           "いちごびす"   "パイン"
[6] "いちごプリン" "YOGURT"        "コーン"        "みつりんご"   "FLAKE"
[11] "ピーナッツ"

$`5`
[1] "冬チョコ"      "カプチーノ"    "ミントチョコ"  "ナッツクランチ"
[5] "ライチ"        "ミックスベリー" "チーズチョコ"  "MILK"
[9] "コーヒーフロート" "WINE"          "レモンケーキ"  "エスプレッソ"
>
```