

EDA of JD data

basic statistic of JD data

user info

id unique

```
## [1] TRUE
```

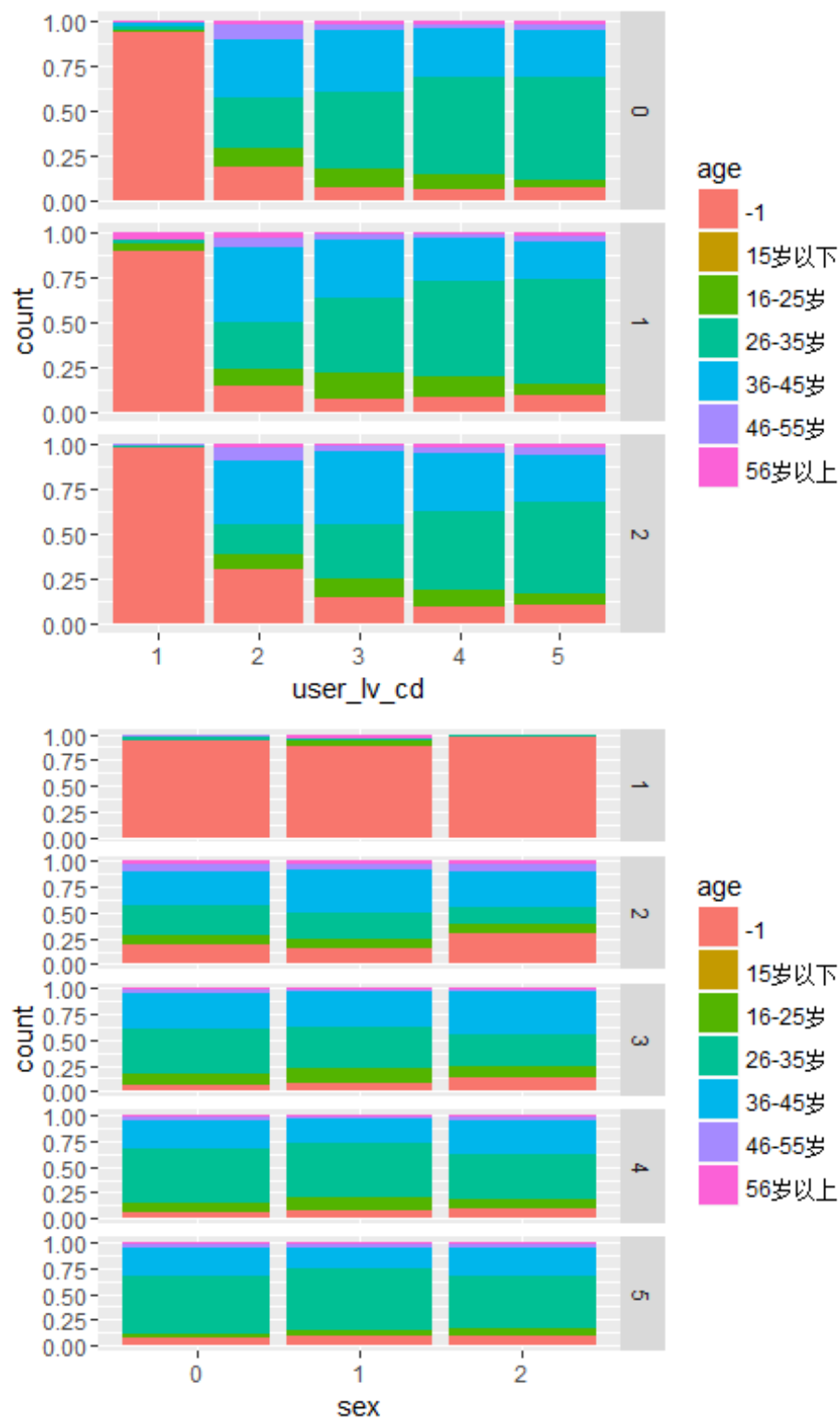
data sample

user_id	age	sex	user_lv_cd	user_reg_dt
54	-1	2	1	2016-01-26
79	36-45 岁	2	2	2016-01-27
100	16-25 岁	2	3	2016-01-27
266	15 岁以下	2	2	2016-01-29
333	-1	2	4	2016-01-30
605	26-35 岁	2	2	2016-02-01

basic summary

```
##      user_id      age      sex      user_lv_cd
##  Min.   :    1   -1      :12803   0:45547   1: 2328
## 1st Qu.:25905  15 岁以下:    6   1: 7585   2: 7519
## Median :51809  16-25 岁 : 7999   2:50484   3:21689
## Mean   :51809  26-35 岁 :46525           4:32205
## 3rd Qu.:77712  36-45 岁 :30828           5:39875
## Max.   :103616  46-55 岁 : 3407
##                      56 岁以上: 2048
##      user_reg_dt
##  Min.   :2003-06-12
## 1st Qu.:2011-11-12
## Median :2013-06-04
## Mean   :2013-04-11
## 3rd Qu.:2015-01-05
## Max.   :2016-11-29
##
```

variable interactive plot



comments info

id unique

```
## [1] FALSE
```

data sample

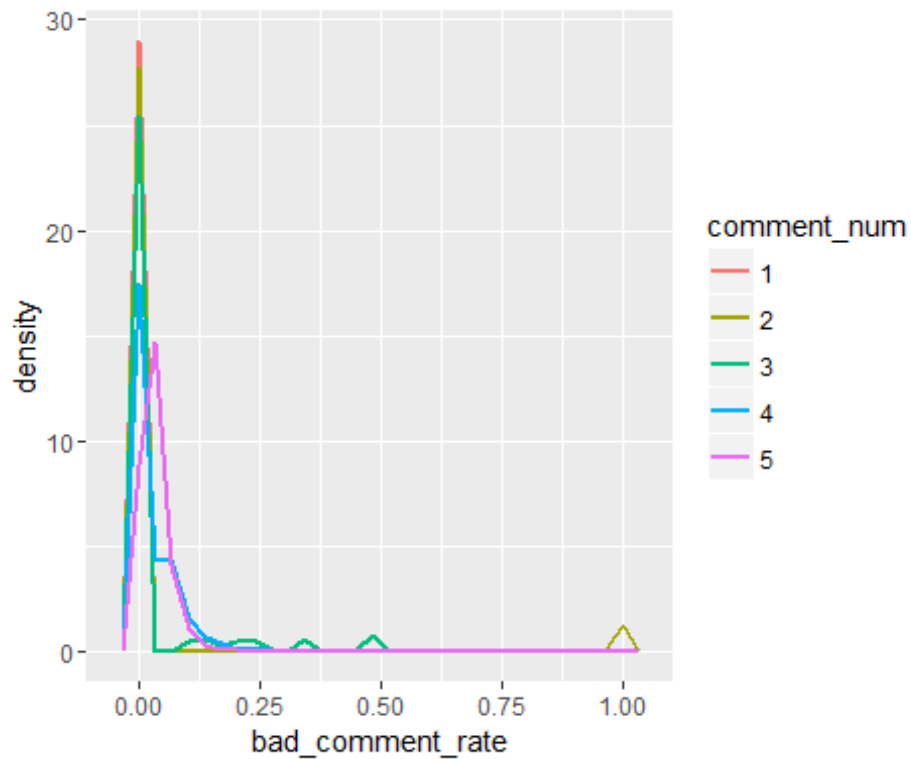
dt	sku_id	comment_num	has_bad_comment	bad_comment_rate
NA	166731	1	0	0
NA	143220	1	0	0
NA	145770	1	0	0
NA	68950	1	0	0
NA	88805	1	0	0
2016-02-01	115070	3	1	0

basic summary

```
##          dt          sku_id      comment_num has_bad_comment
##  Min.    :2016-02-01  Min.    :      1      1:124684      0:124684
## 1st Qu.:2016-02-13  1st Qu.: 43764      2:107403      1:284736
## Median :2016-03-06  Median : 85949      3:117270
## Mean   :2016-03-05  Mean    : 85902      4: 41719
## 3rd Qu.:2016-03-26  3rd Qu.:127914      5: 18344
## Max.    :2016-04-15  Max.    :171230
## NA's    :124684
## bad_comment_rate
## Min.    :0.00000
## 1st Qu.:0.00000
## Median :0.00000
## Mean    :0.02589
## 3rd Qu.:0.00000
## Max.    :1.00000
##
```

variable interactive plot

```
## `stat_bin()` using `bins = 30`. Pick better value with `binwidth`.
```



product info

id unique

```
## [1] TRUE
```

data sample

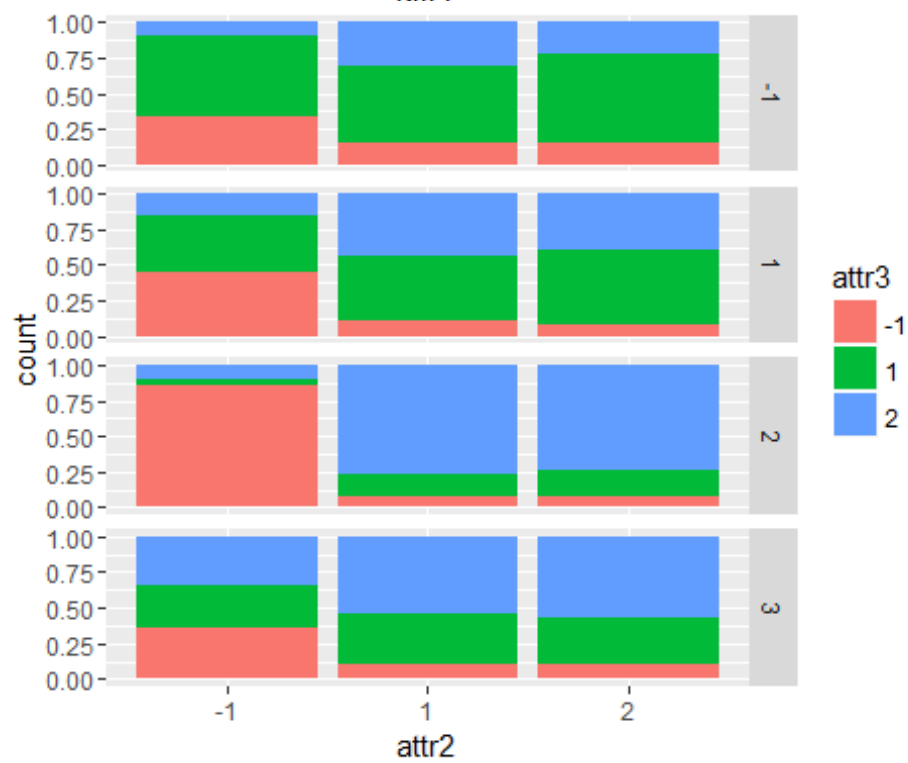
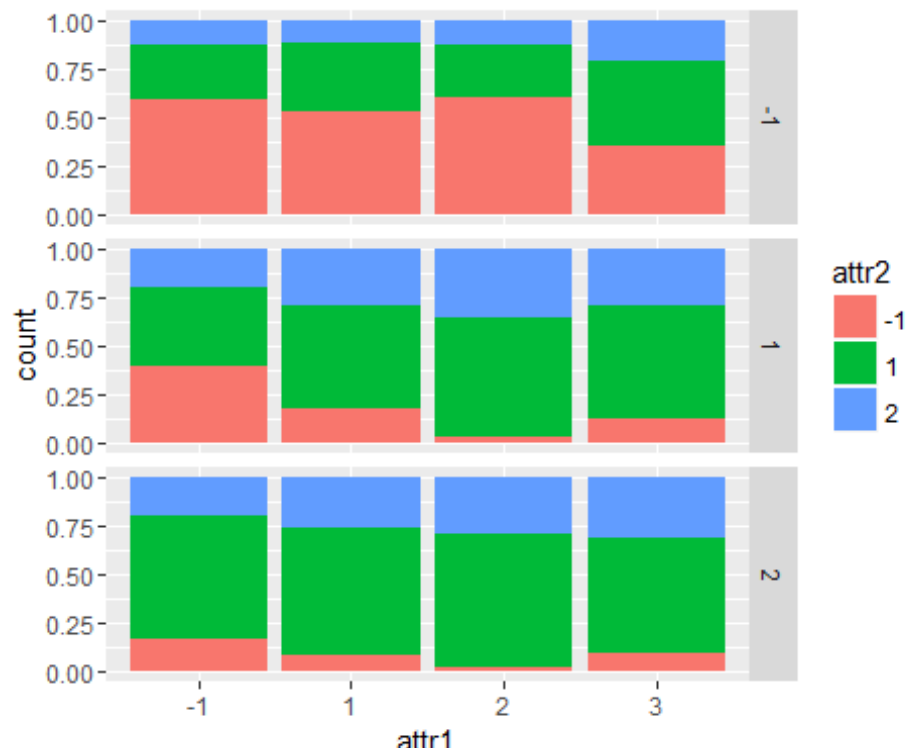
sku_id	attr1	attr2	attr3	cate	brand	popbrand
166731	1	-1	-1	8	545	545
146488	2	2	-1	8	812	812
70546	3	1	1	8	214	214
92487	2	2	-1	8	812	812
113065	2	2	-1	8	812	812
80307	2	2	-1	8	812	812

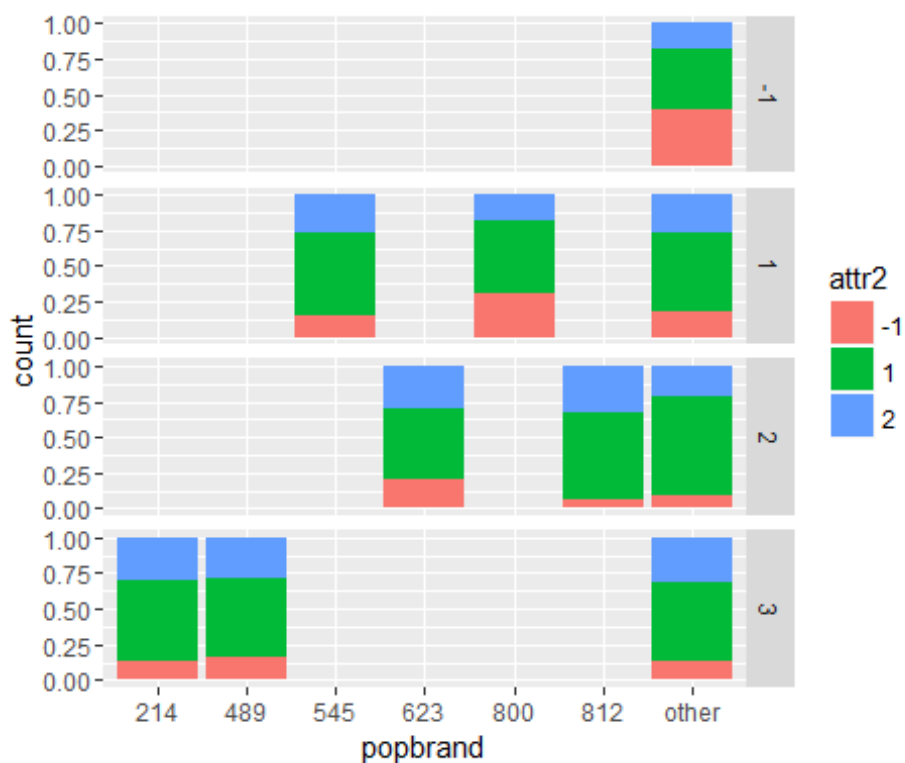
basic summary

```
##      sku_id      attr1      attr2      attr3      cate
##  Min.   :      6    -1: 1701    -1: 4050    -1: 3815    8:24187
## 1st Qu.: 42476     1 : 4760     1 :13513     1 : 8394
## Median : 85616     2 : 3582     2 : 6624     2 :11978
## Mean    : 85399     3 :14144
## 3rd Qu.:127774
## Max.    :171224
##
```

##	brand	popbrand
##	489 :6637	214 :6444
##	214 :6444	489 :6637
##	623 :1101	545 : 945
##	812 :1061	623 :1101
##	800 :1015	800 :1015
##	545 : 945	812 :1061
##	(Other):6984	other:6984

variable interactive plot





act info

str(JData_Action_201603) ###data sample

	user_id	sku_id	time	model_id	type	cate	brand	popbrand
1	70062	168043	2016-03-01	nothit	点击	6	269	other
2	70062	168043	2016-03-01	nothit	点击	6	269	other
3	70062	168043	2016-03-01	nothit	点击	6	269	other
4	70062	168043	2016-03-01	nothit	点击	6	269	other
6	70062	168043	2016-03-01	nothit	点击	6	269	other
7	70062	168043	2016-03-01	nothit	点击	6	269	other

basic summary

##	user_id	sku_id	time	model_id
##	Min. : 1	Min. : 2	Min. :2016-02-29	nothit :2451
488				
##	1st Qu.: 26043	1st Qu.: 41009	1st Qu.:2016-03-02	0 :1874
581				
##	Median : 50260	Median : 79636	Median :2016-03-04	216 : 711
172				
##	Mean : 51274	Mean : 83455	Mean :2016-03-04	217 : 549

413	##	3rd Qu.: 76723	3rd Qu.:126958	3rd Qu.:2016-03-06	218	:	478
449	##	Max. :103616	Max. :171225	Max. :2016-03-08	27	:	242
338	##						(Other):1027
896	##						
##		type	cate	brand		popbrand	
##	点击	:4989971	8	:3005654	214	: 943264	489 : 826
176	##	购物车删除:	44177	4	:1298803	489 : 826176	545 : 3957
75	##	关注	: 21799	6	: 881628	306 : 733816	640 : 204
122	##	加购	: 103680	5	: 743506	545 : 395775	693 : 193
327	##	浏览	:2168509	9	: 687693	800 : 321297	800 : 321
297	##	下单	: 7201	7	: 627958	885 : 302303	885 : 302
303	##		(Other): 90095	(Other):3812706		other:5092	
337							

variable interactive plot

