

# 第六组

## 收视数据分析报告

组员： 李媛媛

陈绍龄

张奕蕾

# 目录

一、	数据背景 .....	1
二、	数据描述 .....	1
2.1	数据概述 .....	1
2.2	异常数据处理.....	2
2.2.1	异常变量的初步处理.....	2
2.2.2	进一步异常分析 .....	3
2.2.3	Session.....	4
2.3	变量描述 .....	5
2.3.1	样本编号 (num) .....	5
2.3.2	收看起始时间 (timestart) .....	7
2.3.3	收看节目 (program) .....	8
2.3.4	收视频道 (channel) .....	13
2.3.5	收看时长 (period) .....	21
三、	数据分析 .....	25
1.	基于 apriori 算法的关联规则分析 .....	25
2.	Session 分析.....	33
3	收视率分析 .....	36
4.	开机情况.....	55

## 一、数据背景

preference.txt 是一个关于收视情况的数据集，采用 UTF-8 编码。其中每一行数据是一个收视样本，包含了样本编号、收视时间、收视频道、收看节目以及收视时长五个基本信息。通过数据集查找，我们将信息来源锁定为 2012 年 2 月初的江苏南京机顶盒记录。我们将通过这个数据集提供的数据来探索与观众收视情况相关的信息。

## 二、数据描述

### 2.1 数据概述

数据概述表

变量名称	样本编号	收看起始时间	收视频道	收看节目	收看时长
代码名称	num	timestart	Channel	program	period
变量类型	整型	h:m:s 时刻变量	字符型	字符型	数值型(秒)

表格 1

注：数据集总共有 2,147,174 个样本。



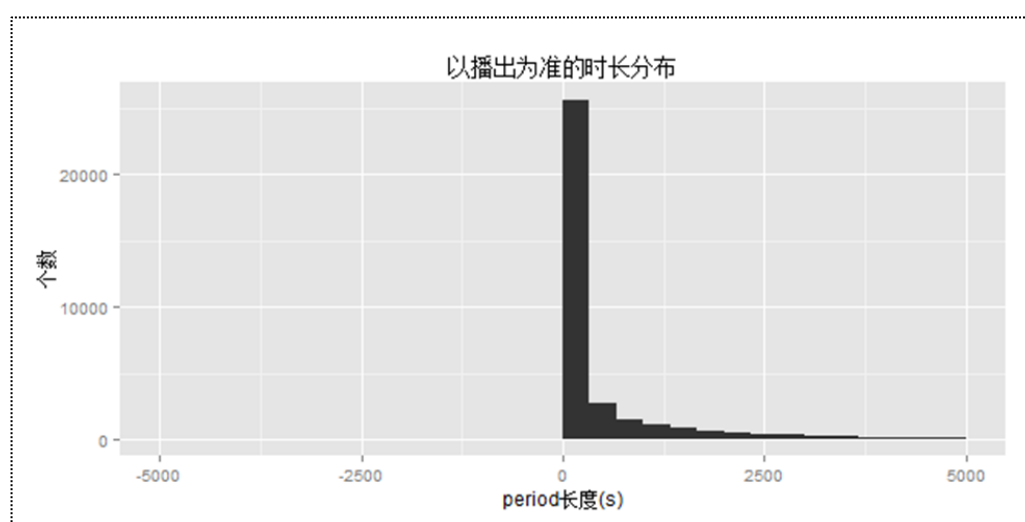
## 2.2.2 进一步异常分析

### 1、 以播出为准的节目变量

#### 1) 分析

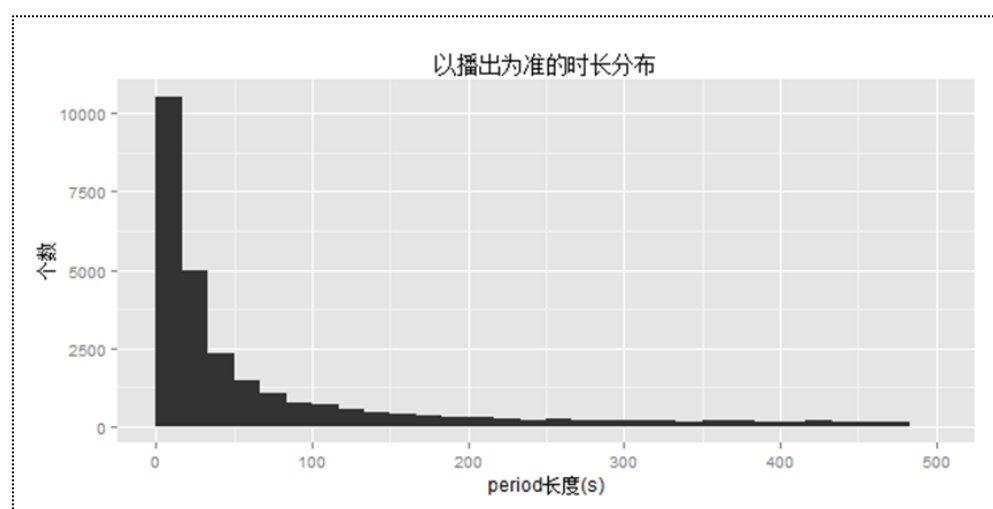
全部样本中共有 35796 个节目为以播出为准的样本。这部分样本所占总时长在所有节目中排名第五，且时长的分布范围很广，但观众并不愿意在这些节目中停留较多的时间（如图 1、图 2）：

总体范围分布在 0~5000 秒：



图表 1

将坐标缩小到 0~500：



图表 2

找出这些节目播出的频道（如表 3），发现这些频道大部分是测试频道、购物频道及生活频道。较能代表节目特征。

节目	频率	节目	频率
快乐购	11541	测试节目 7	277
风尚购物	11292	测试节目 10	267
金陵热播	2401	测试节目 8	258
股市	1592	第一导视	239
测试节目 4	1041	测试节目 9	226
测试节目 5	996	测试节目 13	215
预留频道	983	测试节目 12	200
预留通道	912	测试节目 14	197
测试节目 6	651	测试节目 16	151
测试节目 3	537	测试节目 15	150
测试节目 11	346	纪实频道	14
测试节目 17	338	江宁信息频道	14
数字电视指南	333	江苏卫视高清	1
测试节目 2	320	网络棋牌	1
测试节目 1	303		

表格 3

## 2) 处理

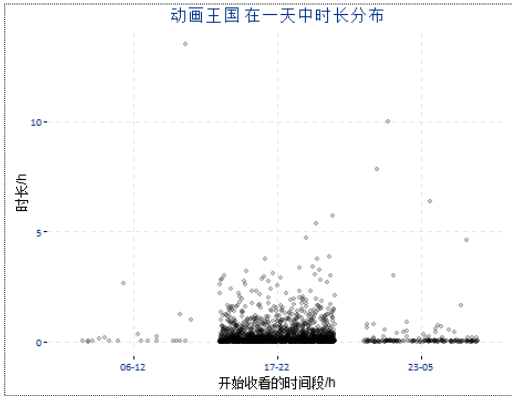
由于这些频道在所有样本中对应的节目只有 NULL 和“以播出为准”，无法用 NULL 填补的方法，故用频道名填补节目名，以减小“以播放为准”在时长、频数分析中所占的比例。

### 2.2.3 Session

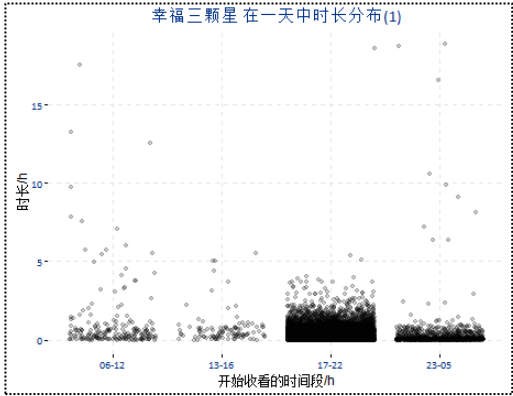
## 1) 分析

观察到时长数据中有很多过大的异常值（可能是由于未关机顶

盒)，并且不同节目的异常分布不同（如图 3），同一节目的不同时段异常分布也不相同（如图 4）。



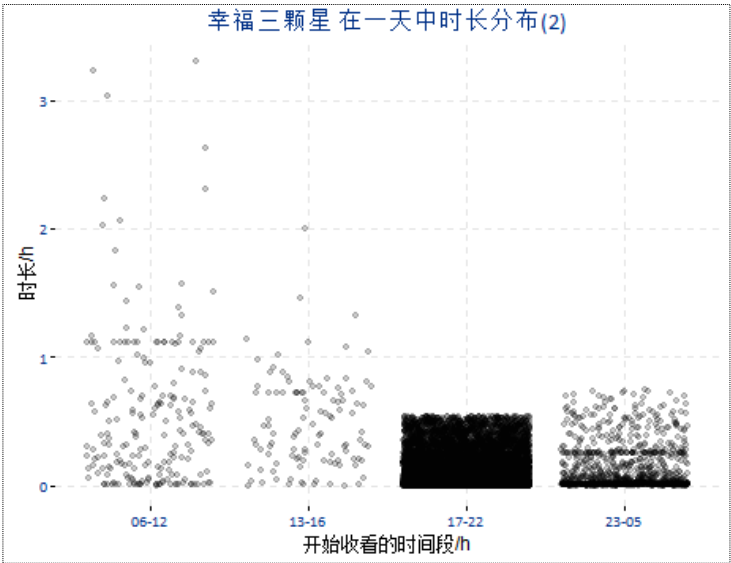
图表 4



图表 3

2) 处理

对每个节目在 6-12、13-16、17-22、23-5 四个时段下的时长取均值，判定 3 倍均值以外的数据为异常。我们对 3 倍均值以外的异常数据设置 session,并用均值填补（以“下午剧场：杨贵妃秘史”为例，填补结果如图 5）。设置后，



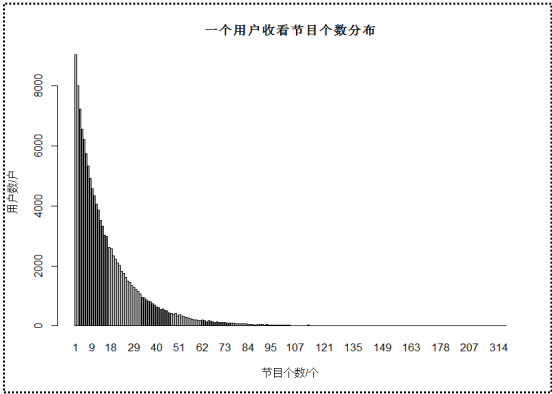
图表 5

平均每条用户一天中有 2.45 次中断（session）

2.3 变量描述

2.3.1 样本编号（num）

清洗后该数据集有 2143493 行，128557 个不同编号，一个编号表示一个用户，一天中一个用户最少看 1 个节目，最多看 447 个节目，平均每个用户一天大约看 17 个节目。



图表 6

由图 6 可以看到大多数用户一天收看节目个数集中分布在 0-100 之间，一天收看节目超过 100 个的用户非常少。

在收看节目较多的用户中，我们查看了编号 825010227049158 的具体信息（如表 4）：该用户一天看 447 个节目，每个节目停留平均时长大约为 137s，可以看到其在一天内不断切换节目，因此我们判断该离群点正常，不作删除。

表格 4

Num	timestart	Channel	Program	period	mark_break
825010227049158	0:36:05	江苏综艺	震撼一条龙精选	6	0
825010227049158	0:36:11	江苏教育	股市直击	10	0
825010227049158	0:36:21	东方卫视	东方直播室	88	0
825010227049158	0:37:49	湖南卫视	相声小品拜大年	68	0
825010227049158	0:38:57	旅游卫视	泰学	28	0
----					
825010227049158	23:26:52	江苏卫视	非常了得	70	0
825010227049158	23:28:02	吉林卫视	欢乐送	8	0
825010227049158	23:28:10	贵州卫视	真相	32	0
825010227049158	23:28:42	大众影院	蝙蝠侠前传之侠影之谜-中 2	33	0
825010227049158	23:29:15	CCTV-2	中国商人	8	1

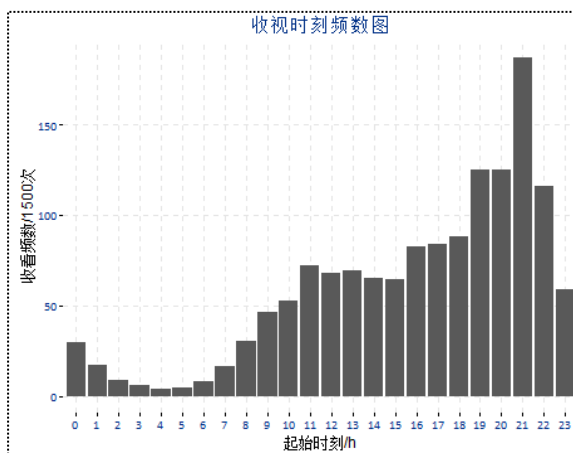


## 2.3.2 收看起始时间 (timestart)

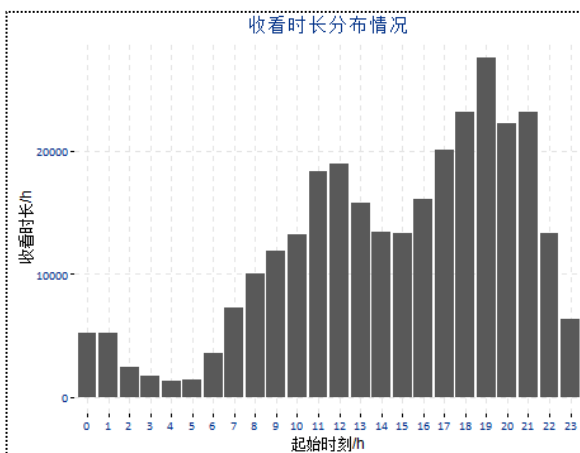
### 1、 整体分析

我们将起始收看时间一小时分割成 24 个时间段。分析人们在各个时间段看电视的喜好。记录起始时间在每个时间段的出现次数，绘制收看时段关于频数的条形图（如图 7），观察到观众在 19：00 左右调频的次数比较高；

再以时间段为划分，记录每个时间段观众收看节目的时长（如图 8），发现 12：00、19:00 这两个时段附近会达到两个小高峰。观众更倾向于在中午 12 点、晚上 7 点左右的休息吃饭时间收看电视节目，晚 9 点开始频繁换台，晚 7 点到 8 点间或存在一些持续时间较长且比较吸引人的节目。



图表 8

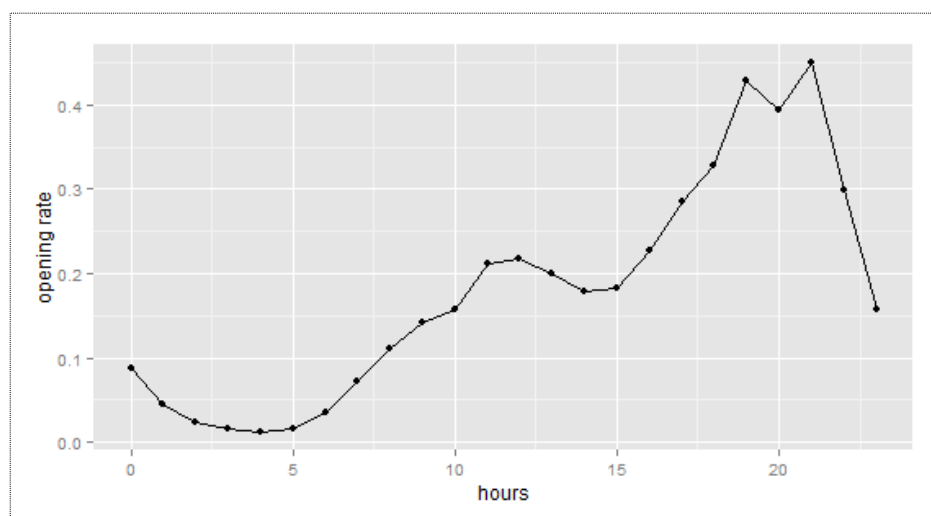


图表 7

### 2、 开机率

开机率是指某一时间段内开机用户占总用户的比例。反映出某一时段内收看人数的多少。我们从零点开始，每一小时划分一个时段，统计该时段内有收视记录的用户数，并除以样本中的用户总数，得到

每一时段内的开机率，画出开机率一时段的折线图如下：



图表 9

由图看出观众的开机率最低是凌晨 3:00 到 6:00，开机率最高的是晚上 18:00 到 22:00，而最高的开机率在 21:00 左右，达到 0.45。接下来我们可以进一步分析，在夜间 18:00 到 22:00 开机的 45% 的观众是否是比较稳定的同一群人。

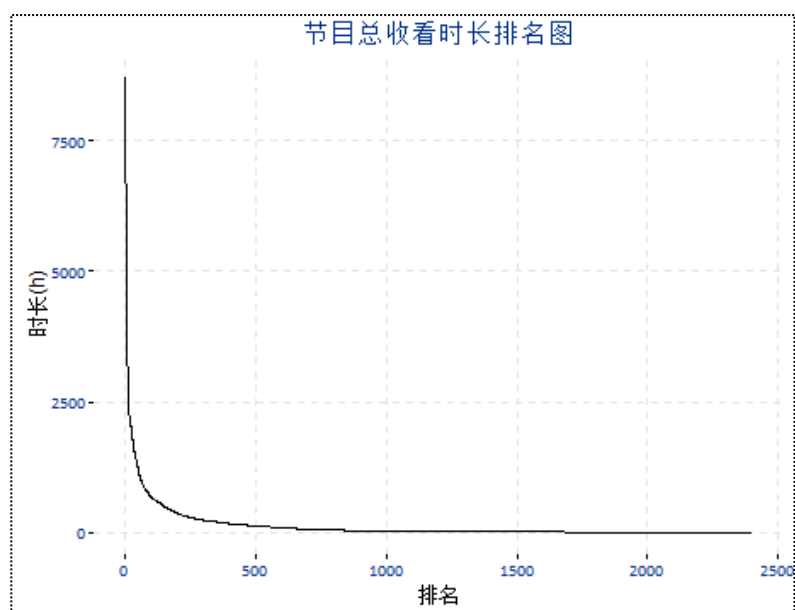
### 2.3.3 收看节目 (program)

#### 1. 节目总数

不对节目文字进行任何处理的情况下，共有节目 3579 个；但是电视剧节目的不同集数应当被视为同一节目，于是在对节目字段做了去除集数处理之后，统计得到的节目共有 2401 个。接下来对节目变量的处理中，我们都将使用去除了集数的变量。

#### 2. 节目与时长的关系

计算所有节目的总收看时长，按从大到小的次序排列，并画出图 10：



图表 10

排名前十的数据如表 5:

	peogram	sumt ime (s)	appearance (次)
1	星光大道-2011 年度总决赛	31364224	39387
2	乡村爱情小夜曲	16730728	37754
3	中国新闻	15774059	18816
4	2012 年春节联欢晚会	13948853	18769
5	零距离	12749517	11369
6	宫锁珠帘	12370912	14805
7	新闻直播间	11882860	15990
8	城市第一剧场: 誓言今生	11158455	17314
9	新闻联播	10825159	14846
10	快乐大本营	9773403	12513
11	J 品赛场: CBA 联赛: 天津一江	9334165	12960
12	中国男子篮球职业联赛第 28 轮	9037865	19710
13	难忘今宵	8763048	11453
14	朝闻天下	8187388	5152
15	2006 年春晚语言节目集锦	7954859	9328

表格 5

从图 10 可以看出, 绝大部分的收视时长都集中在前 500 个节目当中, 而第 500 个节目之后, 每个节目的收视时长几乎都接近于 0。我们计算前 500 个节目的总收视时长以及后 1901 个节目的总收视时长, 对比如表 6:

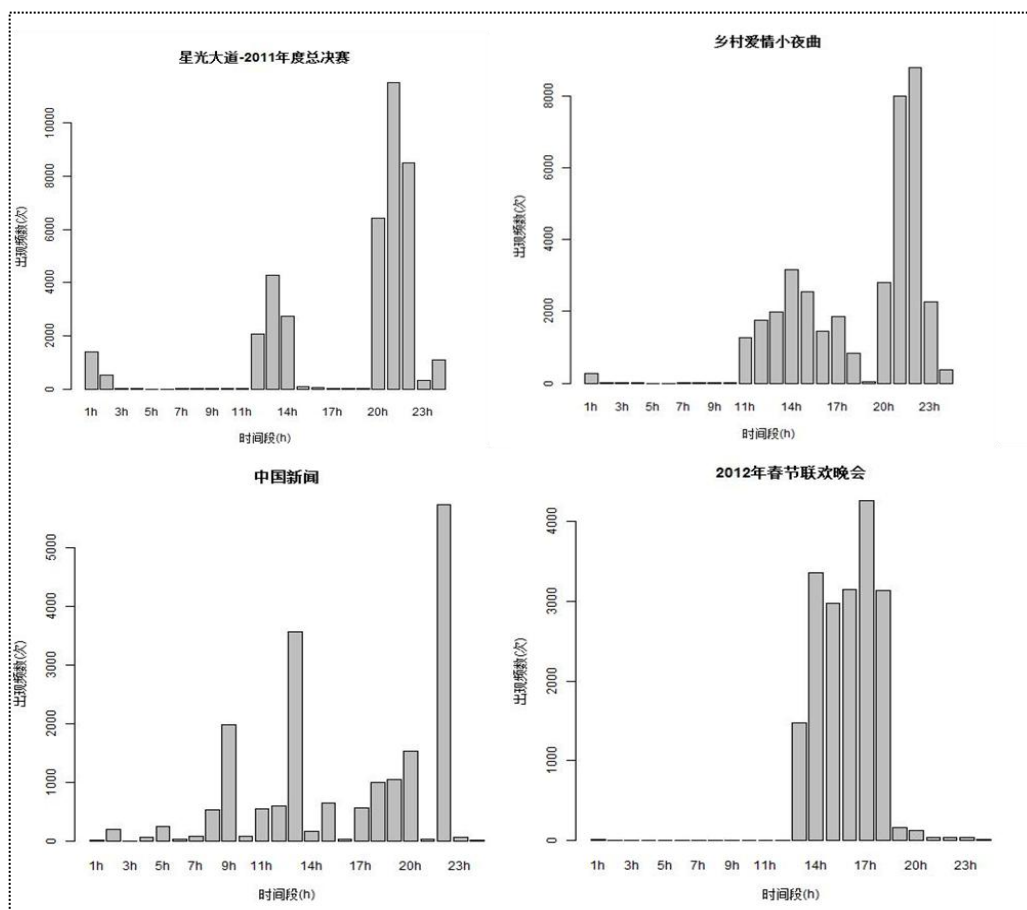
节目	前 500 个	后 1850 个	前 500 个所占比例
总收看时长	569774. 2h	70247. 96h	0. 8902413

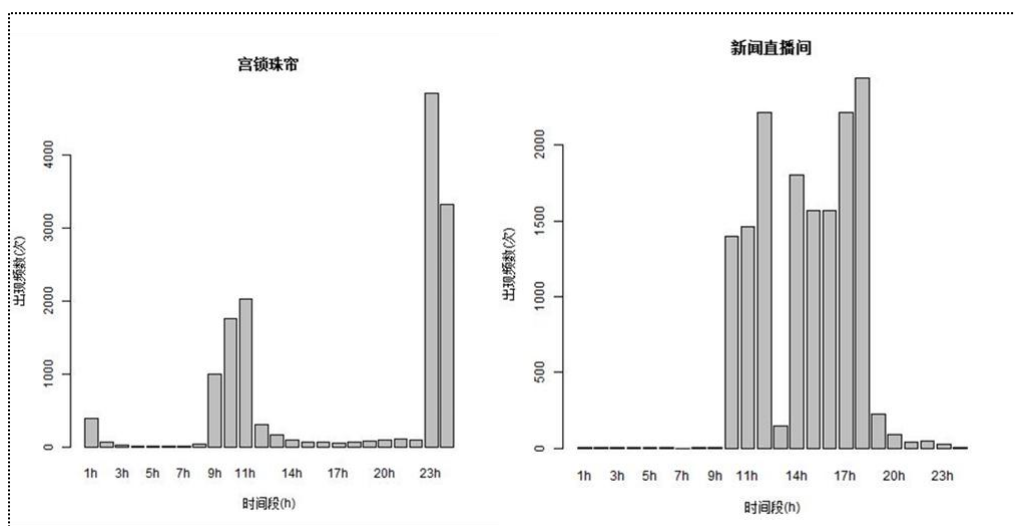
表 6

从上表可以看出，前 500 个节目的总收视时长都占到了全部收视时长的近 90%，这恰好反映出了长尾理论描述的现象：即 20%的热门产品往往可以占据 80%的市场。而我们可以从中得到的信息是：当我们研究什么样的电视节目收视率会高时，可以着重从前 500 个节目下手的；而当我们分析整个电视市场的利润分割时，剩下的 1901 个节目就不能被忽略，虽然这部分中每一个节目的收视率都不高，但总体上仍然占据了 10%的市场，根据长尾效应理论，仍然可以挖掘出很大的市场价值。

### 3. 节目与时段的关系

图表 11





图表 12

根据节目与时长研究中的发现,我们继续研究收视时长排名前五百的节目主要在哪些时段出现,画出排名前八的图(如图 11、图 12):

这些节目都有较为固定的播出时间。普遍来说,大部分观看时间集中在中午和晚上,主要是中午 11:00~15:00,以及 20:00~24:00。新闻类节目的收视时间分布在全天,而综艺及电视剧类节目则主要集中在下午和晚上。

#### 4、节目分类

##### 4.1 对节目按照以下编号分类

类别	节目数量	记录数目	细分	节目数量
新闻 ( 1 )	218	251064	新闻消息 ( 11 )	148
			专题 (12)	44
			谈话 (13)	27
文艺娱乐 ( 2 )	1121	1199542	电视剧 ( 21 )	360

			综艺 ( 22 )	200
			电影 ( 23 )	227
			体育 ( 24 )	152
			其他 ( 25 )	185
教育 ( 3 )	468	34107 1	社教 ( 31 )	102
			少儿 ( 32 )	173
			其他 ( 33 )	191
服务 ( 4 )	569	34514 2	生活服务 ( 41 )	384
			理财 ( 42 )	75
			广告 ( 43 )	77
			其他 ( 44 )	31
无法分类	25	6674	无	无

表 6

备注：无法分类的节目包括：彩条、重播昨晚、结束语、公视时段、开始语、全天节目结束、技术维护检修、垫播等。

#### 4.2 频道和节目分类点图

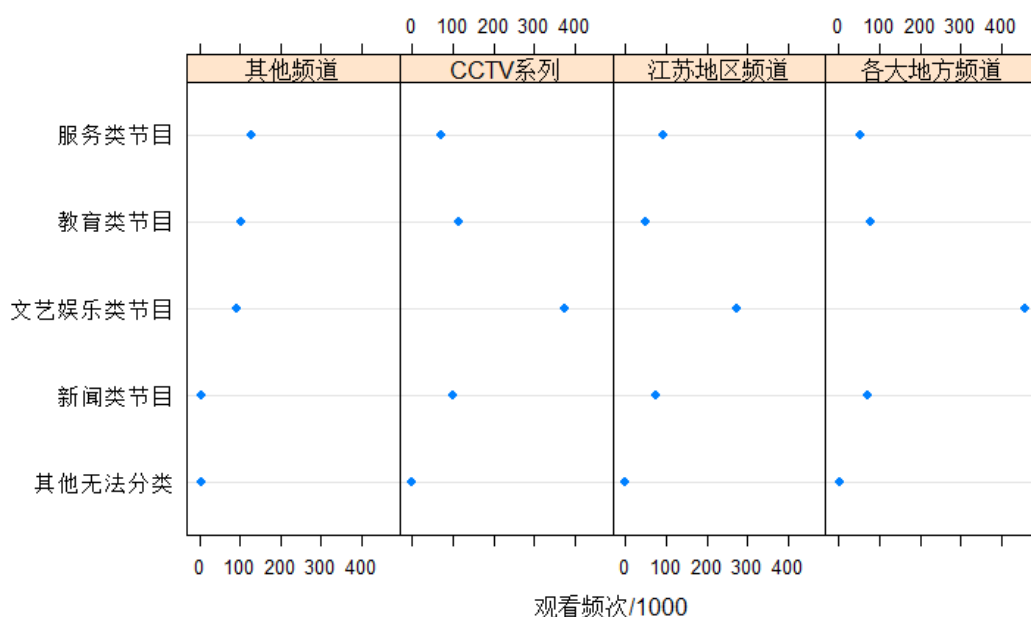


图 13

从横向看，CCTV 系列的教育类节目稍多于其他频道，各大地方频道的文艺娱乐类节目比重比较大；从纵向看，不同频道之间各自节目的比重分布基本一致，文艺娱乐类节目较多，其他类型的节目比例较少。

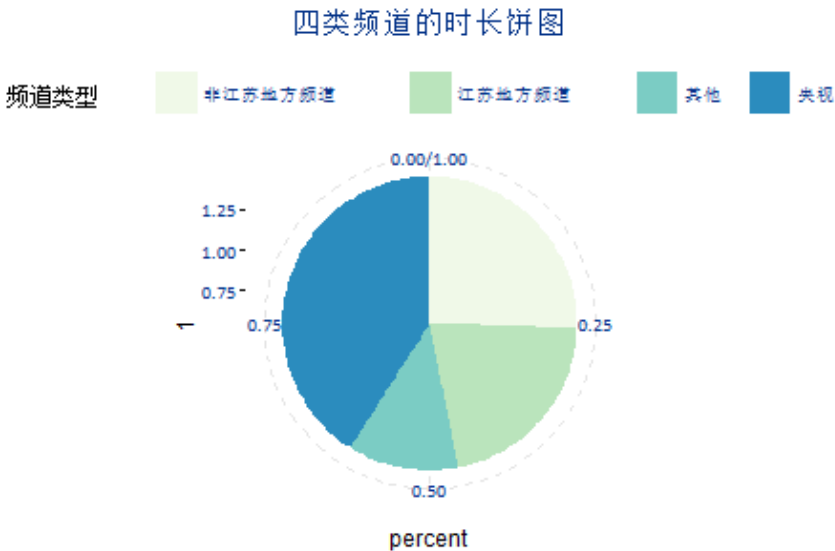
#### 4.3 变量两两关系（下周）

以用户编号为划分，探究各个用户的收视总时长、各类节目或频道比例、收看高峰时间、换台频率等变量的两两关系。

#### 2.3.4 收视频道（channel）

删除无效数据后，我们一共得到了 178 个频道。欲分析观众对电视频道的偏好程度，我们选取了观看总时长达排名前 10 的电视频道作饼状图（图 13），可以看到 CCTV-3、CCTV-1、CCTV-4、江苏城市等频道的观看时长在所有频道中占到了较高比例。

相对于地方频道，央视频道收看的总时长普遍较高，这是由于北方地区，机顶盒设置一般从央视进入。另外发现电影、音乐频道，新闻频道，综艺频道等更受观众喜爱。



图表 14

相对于地方频道，央视频道收看的总时长普遍较高，这是由于北方地区，机顶盒设置一般从央视进入。另外发现电影、音乐频道，新闻频道，综艺频道等更受观众喜爱。

1. 频道与节目的关系

对节目出现的所有频道进行总结，并按照节目出现的频道数进行排序。所有节目中出现频道最多的节目是转播中央台新闻联播，而全部 2341 个节目中有 1743 个节目只在一个频道播出，这些节目占比为 74.45%。播出频道最多的前 16 个节目如下：

1	转播中央台新闻联播	9	刁蛮俏御医
2	新闻联播	10	乡村爱情小夜曲
3	天气预报	11	西游记
4	2012 年春节联欢晚会	12	剧场
5	结束	13	晚间新闻
6	午夜剧场	14	幸福三颗星



7	星光大道-2011 年度总决赛	15	爱情睡醒了
8	宫锁珠帘	16	阳光路上

表格 6

与 2.3.3 中的热门节目对比，可以发现节目有较大的重叠。说明大部分收看时长的节目在很多电视台同时播出，这些节目也占据了大部分收视份额。近 75% 的节目都只能在一个电视台播出，但这些节目总体仍然占据了不小的收视份额。

分别统计每个频道中出现的节目，出现节目最多的频道是零频道宣传，总计 138 个节目；所有频道出现节目数的中位数是 15 个，平均数是 17.6 个。而出现节目数最多的十个频道如下：

排名	频道	节目数	排名	频道	节目数
1	零频道宣传	138	6	南京新闻	47
2	大众影院	69	7	CCTV-3	44
3	CCTV-6	52	8	卫生健康	42
4	CCTV-5	49	9	CCTV-1	41
5	南京影视	48	10	南京少儿	41

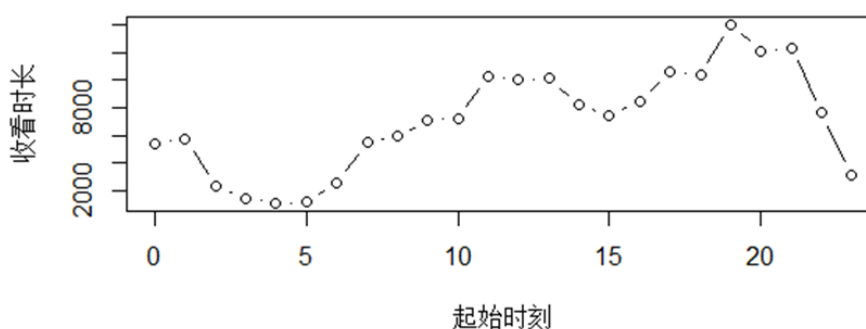
表 8

经过查询资料可知，记录到节目数最多的零频道宣传是 2012 年江苏有线公司在云媒体电视中新增的帮助平台。用户点击该平台可以查看业务介绍、操作指南以及最新的优惠活动。也可以查看视频简介。

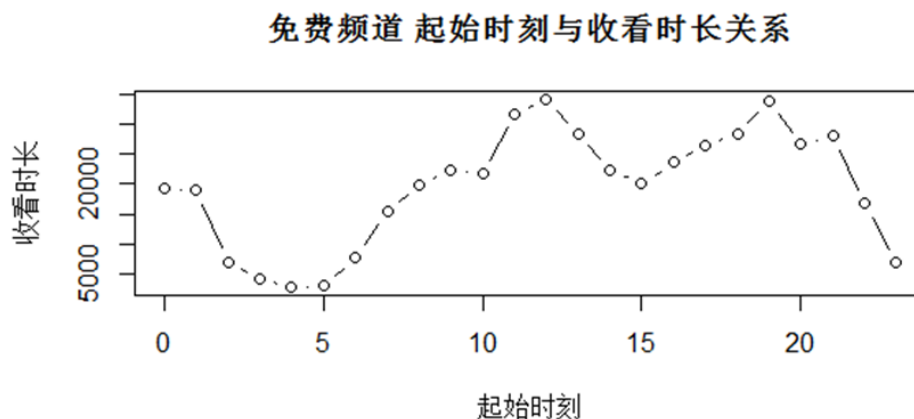
## 2、收费频道与免费频道对比

我们提取 54 个收费频道（一共有 178 个频道），分析收费频道和公共频道之间的差别（图 14、15）。

收费频道 起始时刻与收看时长关系



图表 13

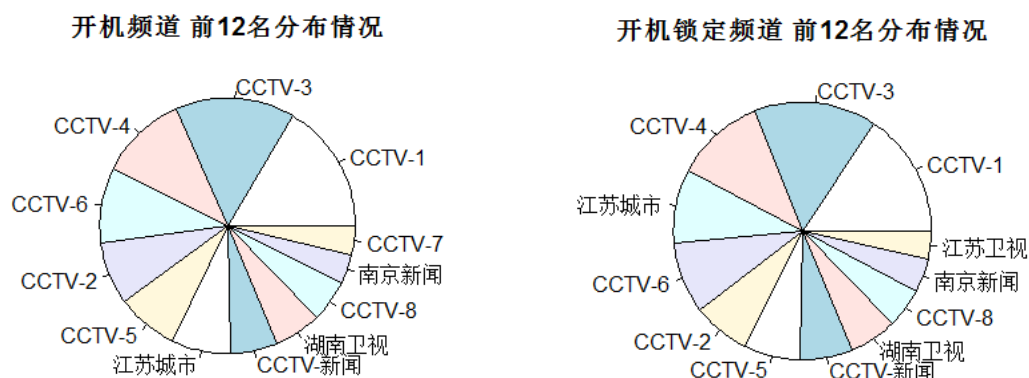


图表 14

由上图看出,收费频道和免费频道的收看时长随时间变化的变化趋势相似,但是 10-15 点免费频道的变化波动明显大于收费频道的变化,主要原因是这段时间并非电视节目播放的黄金时间,许多收费频道在这段时间播出的节目对观众的吸引力会下降。

### 3、开机锁定频道分布

由图 16 可以看出不少用户开机频道为 CCTV 系列,南京新闻和江苏卫视两大地方频道也占据不小比例。我们将开机后观看时长  $\text{period} < 300\text{s}$  的换台数据去除,得到开机锁定频道的分布(如图 17)。

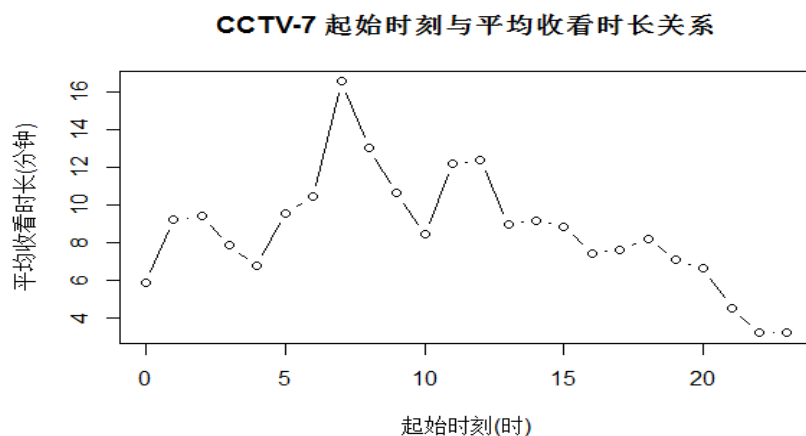


图表 15

图表 16

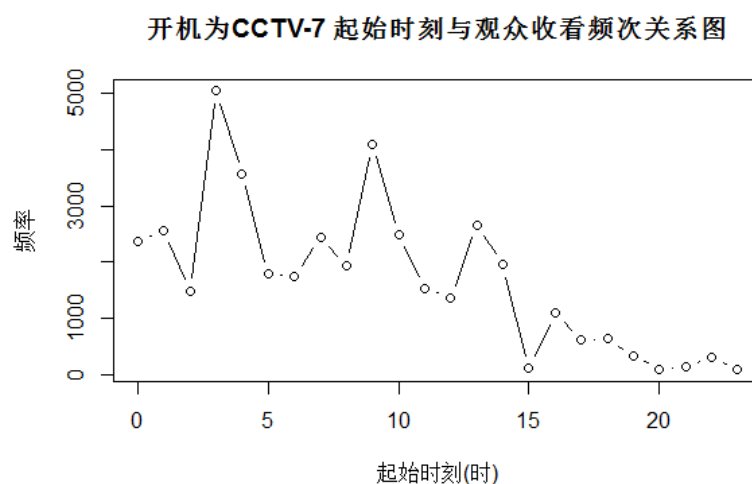
频道分布基本没有变化，只少了 CCTV-7，新增江苏卫视。

我们试着分析 CCTV-7 不在开机锁定频道的原因。



图表 17

可以看出 CCTV-7 在 0-10 点的观众平均收看时长相对较长，可能是这段时间该频道播出的节目比较吸引观众或者由于这段时间可选的频道节目不是很多。这也说明了为什么第二天的开机频道很大部分是由于前天的观看记录造成的。

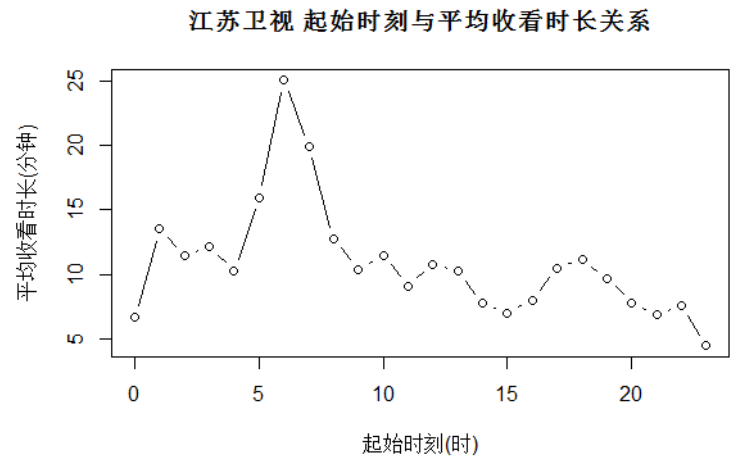


图表 18

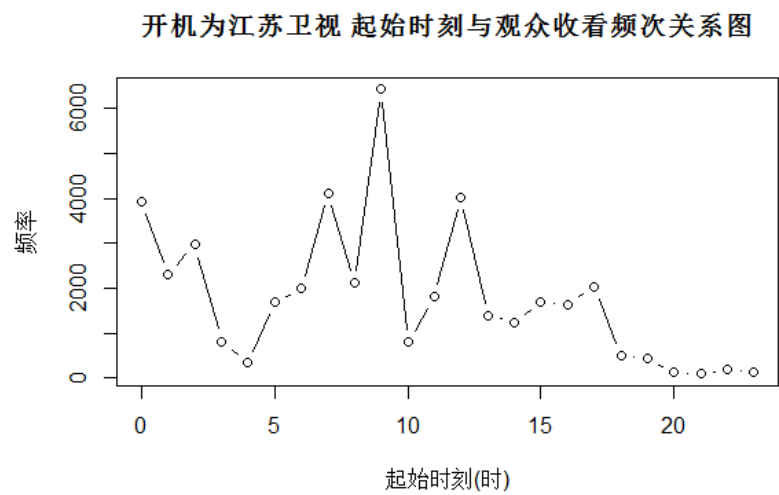
可以看到开机为 CCTV-7 的情况下，不同开机时刻出现的频次图与上图几乎相反，观众很有可能会在 16 点开机，而此时正好是该频

道的非热门节目播出时间，所以造成了不少观众换台。

下面分析江苏卫视。



图表 19



图表 20

与 CCTV-7 相同，观众收看江苏卫视多集中在 5-10 点。但是开机频道为江苏卫视的情况下，频次最多的落在 5-10 点之间，刚好该频道相对热门节目播放，所以没有遭遇观众的频繁开机换台。

#### 4、广播频道

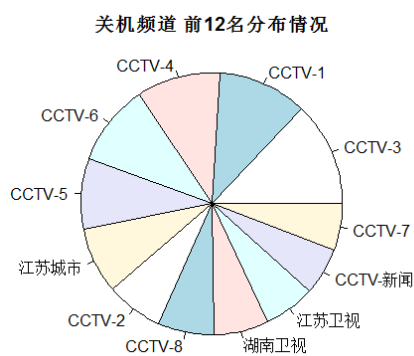
江苏音乐广播	江苏经典流行音乐	金陵之声【FM99.7】	江苏新闻综合广播
--------	----------	--------------	----------

【FM89.7】	【FM97.5】		【AM702】
南京六合音乐 【FM103.5】	江苏交通广播网 【FM101.1】	江苏财经广播 【FM95.2】	国际广播电台 【FM92.4】
南京新闻广播 【AM1008】	南京城市管理广播 【AM1170】	南京交通广播 【FM102.4】	中华之声
江苏健康广播 【FM100.5】	南京音乐广播 【FM105.8】	中国之声【AM1359】	江宁之声【FM88.5】
音乐之声【FM98.9】	江苏新闻广播 【FM93.7】	江苏文艺广播 【FM91.4】	江苏故事广播 【AM1206】
南京新闻广播 【FM106.9】	南京经济广播 【AM900】	南京体育广播 【FM104.3】	南京城市调频 【FM101.7】
华夏之声	民族之声	戏曲曲艺广播 【AM801】	经济之声 【FM107.5】

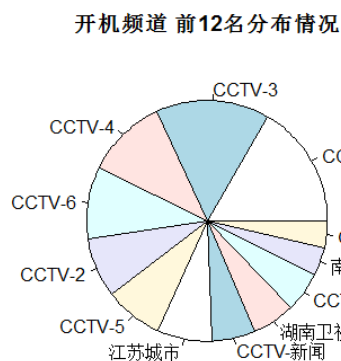
表 9

注：AM（Amplitude Modulation）调幅广播，一般中波广播使用  
FM（Frequency Modulation）调频广播

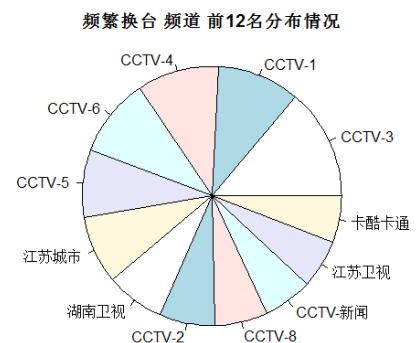
## 5、不同频道指标下频道排名



图表 22



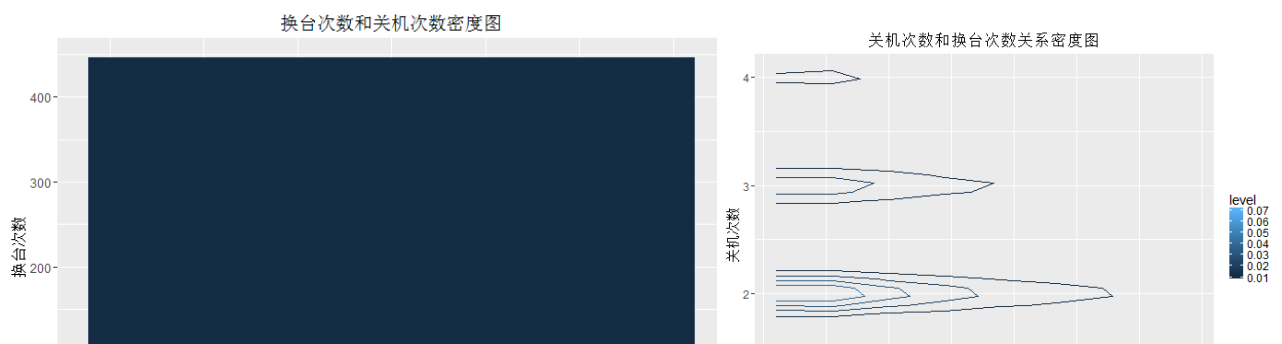
图表 23



图表 24

以上分别是关机次数排名、换台次数、开机锁定次数排名最前 12 位的频道分布，可以看到基本没有什么变化，说明了这些指标并不能单独反映频道是否热门或者冷门。

## 6、换台次数和关机次数关系



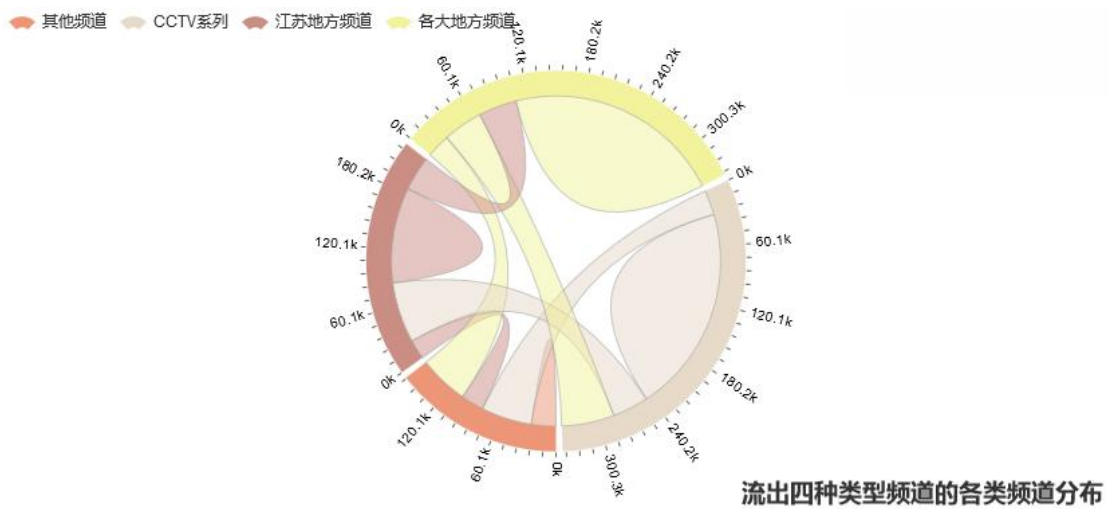
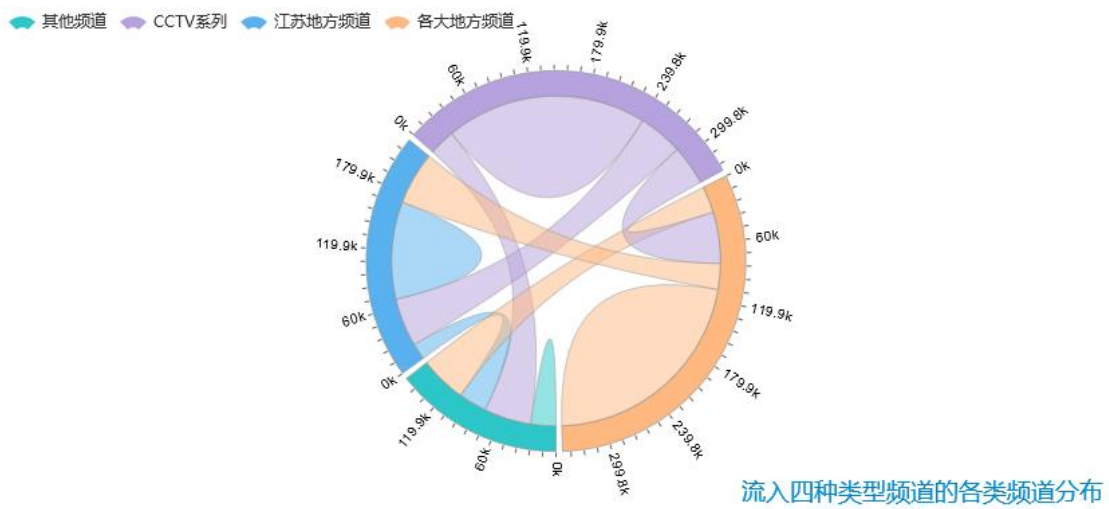
图表 25

图表 25

可以看到关机次数和换台次数都分布在比较低的区域，关机次数多的换台次数相对更少。

7、 频道分类

我们将频道分为四类：其他频道（0）、CCTV 系列（1）、江苏地区频道（2）、各大地方频道（3）；画出各类频道流入流出情况。



可以看到 CCTV 和各大地方频道的比重较大，且这两种频道的内部流动占据主体。

### 2.3.5 收看时长 (period)

#### 1. 观看与起始时间差：

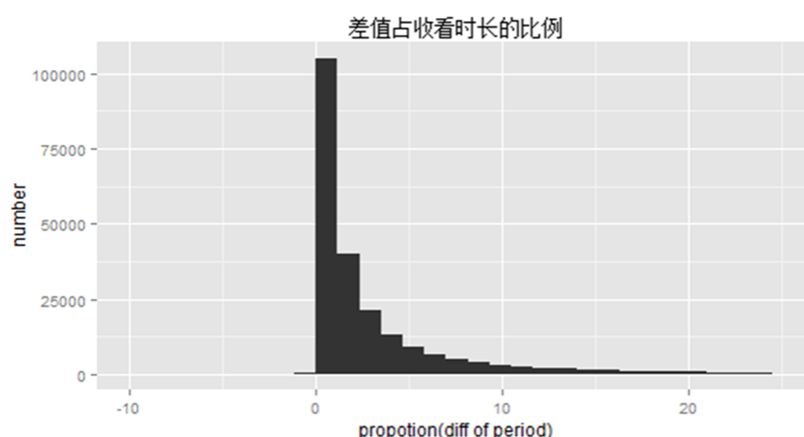
起始时间差是指相邻两个节目开始变量 (timestart) 的差，而时长是指数据中给出的时长变量 (period)。一直以来我们的猜测是时长是准确的收看时长，而两者差是两次记录到电视节目的时刻，我们认为两者不相等的数据则可说明发生了关机事件。但是根据老师所说的数据收集的方法，这两者应该相等，但是我们也观察到这两者其实有一定的差别，于是需要进一步考察这两者的关系。

#### 2. 基本情况

全部样本中共有 361198 个值不相等，占到样本总数的 16.822%，但考虑到跨天数据中的起始时间差和时长本身就不是相等的，我们除去这部分信息再做分析。则一共需要对 2018564 个样本的起始时间差和时长进行分析。

#### 3. 两者相差值占时长 (period) 的比例

为了衡量差值的大小，我们求出两者差值占时长数据 (period) 的比例，然后画出比例分布 (如图 22)：

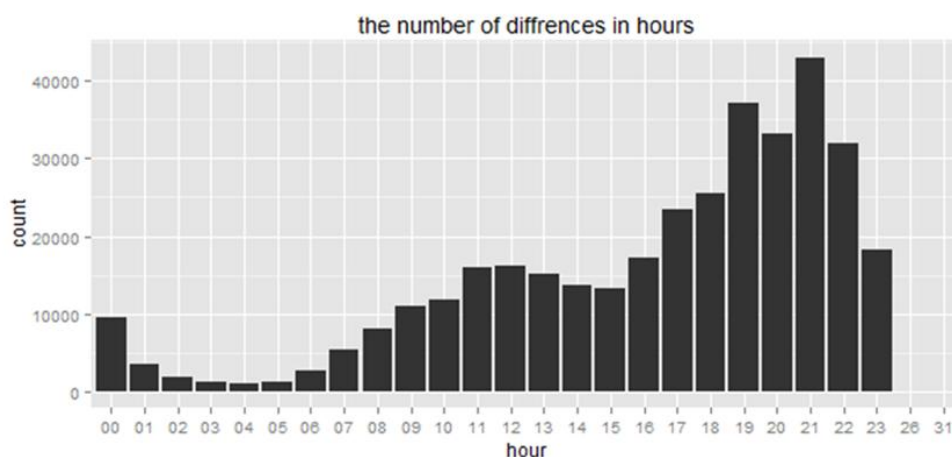


图表 26

从图中可以看出，在两者不相等的的所有数据中，两者差值占时长的比例都浮动范围很大，从 0 到 20 左右皆有分布。就是说这个差值是一个非常随机的值。这与老师所说的数据集的收集方式有一些矛盾，因为有相差的数据很多，并且差值非常不稳定。

#### 4. 差异数据与时间的关系

接下来我们分析有差异的数据个数与时间的关系，我们画出数据按照小时分布的图像如下（这里没有将两个时刻变量记录错误的数据删除）：



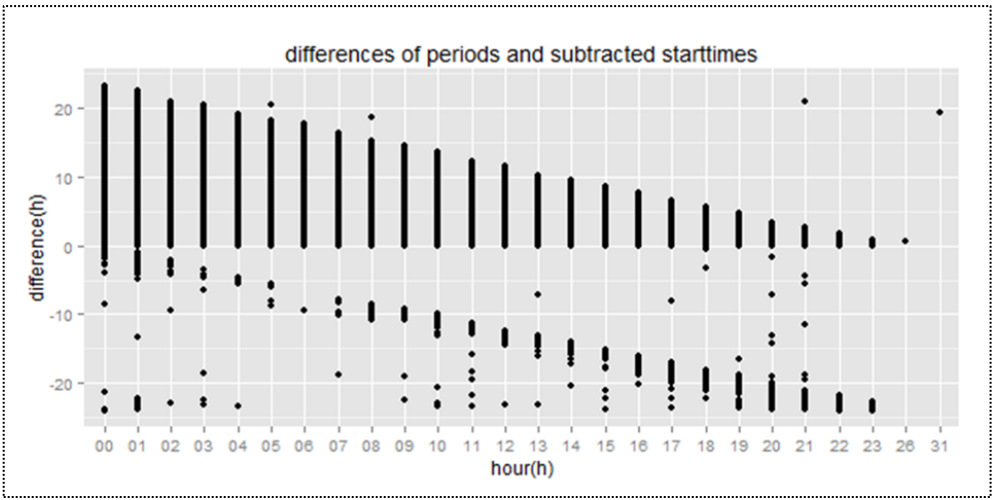
图表 27

这一图像与之前作出的收视频数图对比可以发现，除了在 19:00 时差异数据的比例比较高，其他时间段的数据图像非常相似，说明这种不相等的情况是按一定的比例在任何时段收集的数据中发生。从这一角度来看，似乎我们之前所认为的假设，即：认为时间差是关机时



间，似乎不太吻合，因为人群在每一个时间段内关机的时间应该不是稳定的。而这可能意味着这两者会以一定的比例产生误差。可能是某种特殊原因造成的稳定差异。而在 19:00~20:00 这个时间段内，发生两者不等的概率更大。

接下来，我们对这些差异数据的大小做进一步的研究，我们画出



各时段内，差异数据的时长分布，发现了非常有趣的规律：

图表 28

首先，从这个图中可以看出有很多差值为负，即起始时间差小于 period 值，也就否定了我们之前的想法，即认为时间差代表关机时间。

而从这个图像中，我们看出，从 0:00 到 24:00，正值数据中，两者相差的时长是在以一个稳定的斜率下降的，而负值数据中，两者相差的时长绝对值以相同的斜率增加，表现上是以相同斜率下降。

再观察它的斜率，基本接近于-1。而我们又想到，随着小时数往后推移，下一条记录记录到比现在的样本时间记录大的概率以斜率-1 均匀下降，而比现在的样本时间小的概率则逐渐增加。也就是说

timestart 的差值与误差也有相同的变化规律。这说明,误差的范围,跟 timestart 差值的范围是一样的,结合之前所看到的误差范围非常大,以及每个时段误差比例大致相同,我们可以发现其中的一些规律:

每个时段内的数据都会以一定比例产生这种误差,而误差范围与真实的起始时间差相近,也就是说现在得到 period 很可能是在得到了准确的起始时间差后,部分数据错误的加上或减去了相近位置的其他时间差。

我们对此想法做了一些验证,尝试将起始时间差相邻位置的数据相加减,得到的结果与 period 相比较,发现确实会有一些相同。考虑到这种错误的随机性,我们无法穷举出错原因的所有可能性。但在尝试的这几种结果中,都可以看出我们的猜想是有较大可能性的。

## 5. 结论

(1) period 应该确实来自于起始时间的相减,但是在操作过程中出现了一定的误差;

(2) 出现这种误差的概率在一个用户不同时段的记录中出现的概率相等。即可能在一个用户所有记录的任意行数以近似的概率出现;

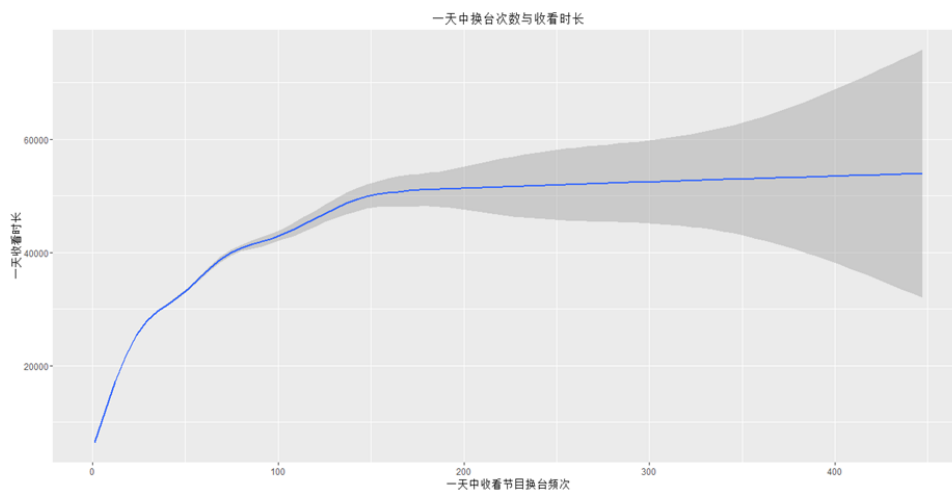
(3) 这种误差可能来源于对于原始的时间差数据无意中进行了相近位置的加减运算。

(4) 不应该根据 period 变量来判断节目收看长短,而过长的 period 往往意味着可能发生了关机行为。

## 6. 一天中换台次数和收看时长的关系

我们可以看到在一天中换台频繁的情况(图 29),相对应的一天的总

收看时长也更长：



在低频次换台的情况下，收看时长随着换台次数的增加而增加；而高频次换台的情形，收看时长基本不变。

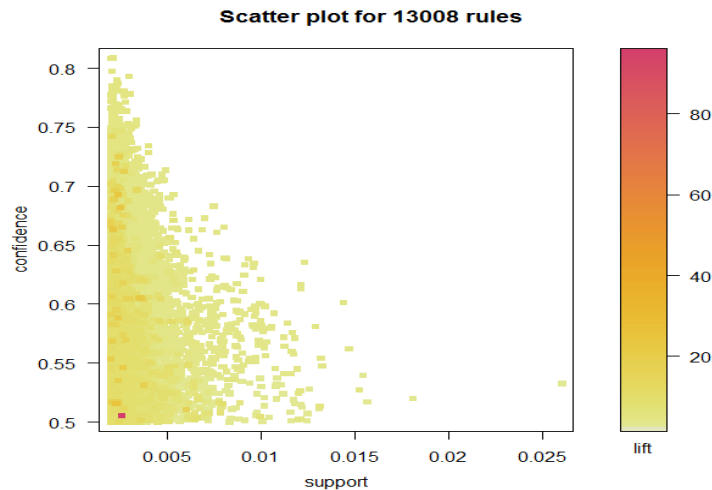
### 三、数据分析

#### 1. 基于 apriori 算法的关联规则分析

我们尝试通过分析不同节目同时出现在同一个编号的潜在关系，试图探究观众对节目之间是否存在某种偏好，比如大多数喜欢 A 节目的观众也会喜欢看 B 节目。

这里我们设置支持度=0.002，置信度=0.5（经过对支持度和置信度的多次调整，对比提升度确定阈值），得到 13008 条关联规则。

##### 1.1 直接绘制关联规则散点图



图表 30

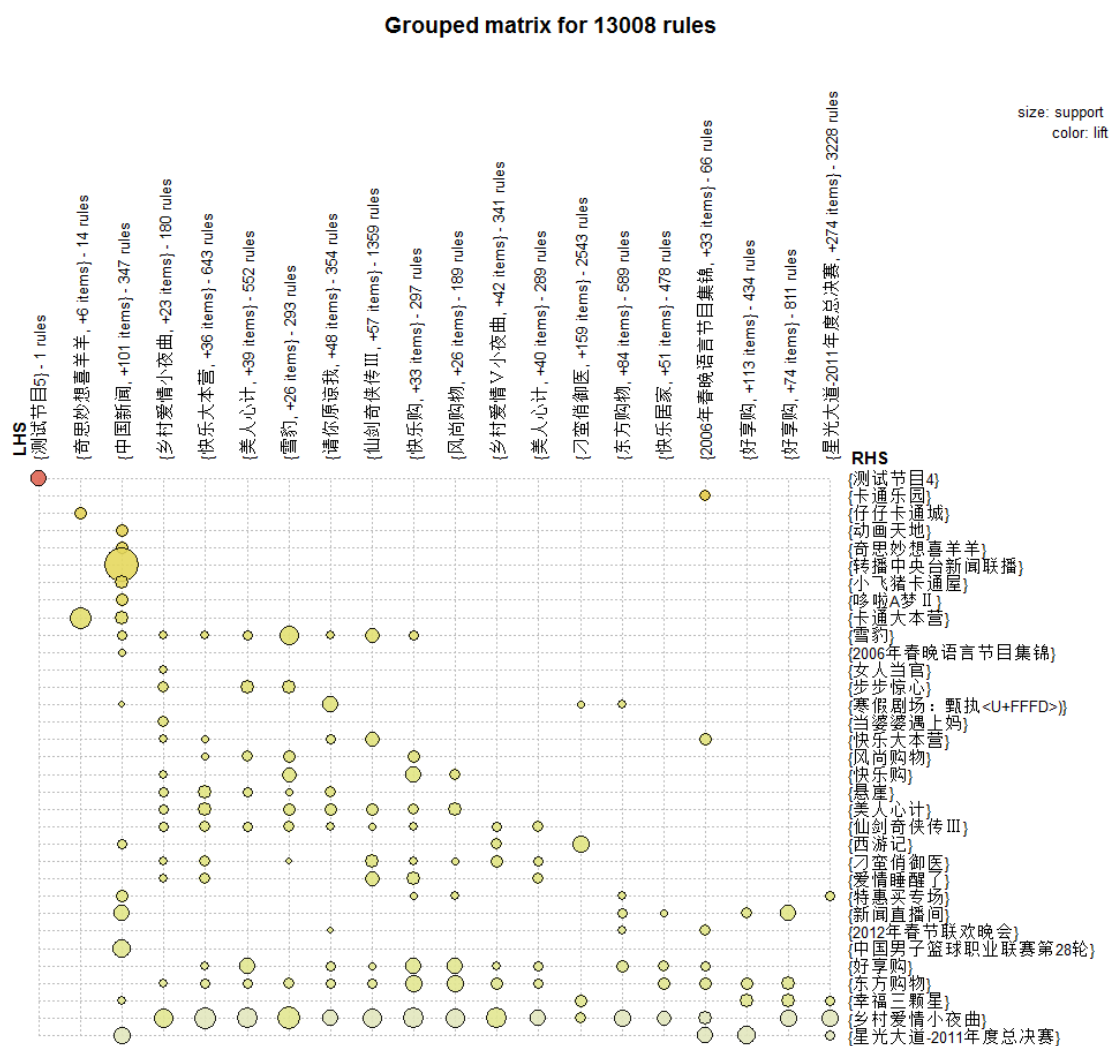
如图 30 可以看出支持度大多数分布在 0.001-0.005，支持度和提升度成反比。分析关联规则散点图中的离群点：支持度很高但置信度和提升度。

rules	support	confidence	lift
{乡村爱情V小夜曲} => {乡村爱情小夜曲}	0.026050701	0.53277124	2.857025

乡村爱情 V 小夜曲播出频道为“山东卫视”，乡村爱情小夜曲播出频道为黑龙江卫视、北京卫视、天津卫视、辽宁卫视。

乡村爱情 V 小夜曲和乡村爱情小夜曲同时出现的概率很高，但两者之间并没有明显的关联关系，说明支持度高是由于两者的播出时间、频率高造成的。

## 1.2 基于分组矩阵的可视化



## 1.3 基于图的可视化

### 1.3.1 支持度排名前 100 的关联规则分布

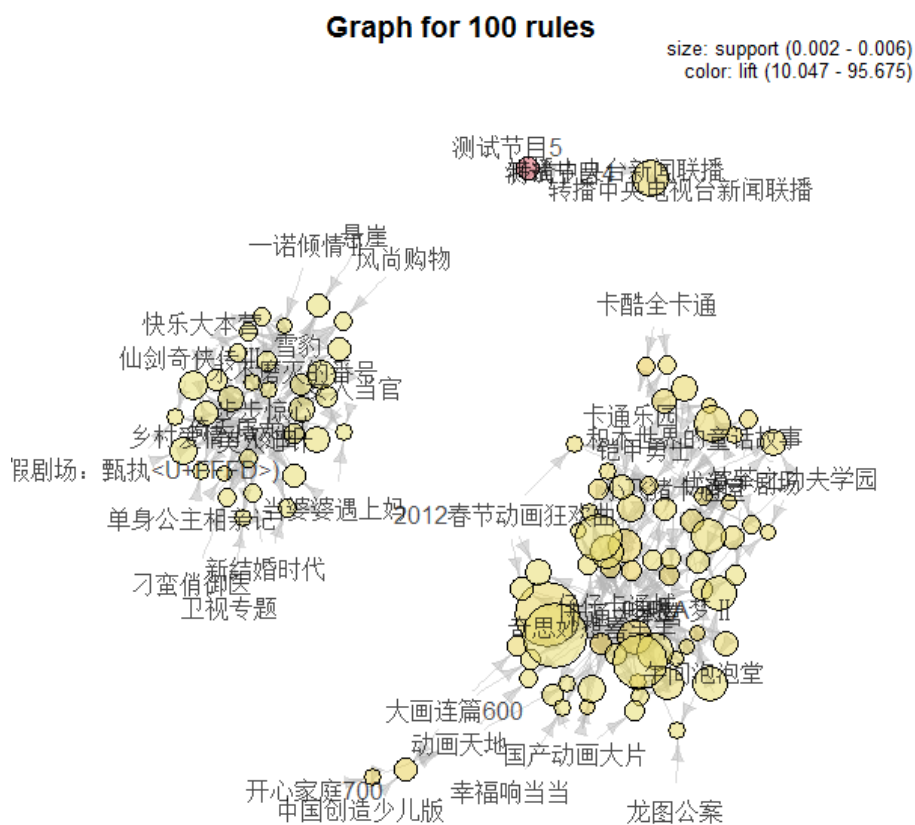


图 32

由上图我们可以大致看出支持度排名前 100 的关联规则分布，主要有三类：少儿节目类、新闻类、连续剧类，我们具体分别进行分析。

### 1.3.2. 新闻类

分析数据集，有 55 个频道转播中央台新闻联播，6 个频道转播中央电视台新闻联播，转播时间与直播时间相同，为 19:00-19:30。我们将这 61 个频道进行统计，得到转播新闻联播节目在这些频道的总收视时长前 10 名（备注：CCTV-1 直播新闻联播的总收看时长为 12531455s）。

将下表与各个频道所有节目总收视时长的排名对比，我们发现几乎一致。说明了同一档节目在不同频道的收视情况也会有所不同，同

一节目在越热门的频道得到的关注度也要更高。

排名	频道	收看频次	收看新闻联播转播 总时长	排名	频道	收看频次	收看新闻联播转播 总时长
1	湖南卫视	758	383006	6	安徽卫视	428	132729
2	北京卫视	582	227503	7	湖北卫视	425	113363
3	天津卫视	528	173590	8	山东卫视	403	134192
4	浙江卫视	505	179720	9	四川卫视	363	94368
5	辽宁卫视	443	142058	10	吉林卫视	352	98940

表 10

此外，我们可以看到“转播中央台新闻联播”“转播中央电视台新闻联播”为同节目（只是名字上差异），所以在后续的操作中，我们将数据中的这两类节目统一命名。

1.3.3. 支持度排名前十的规则分布（主要为少儿类）

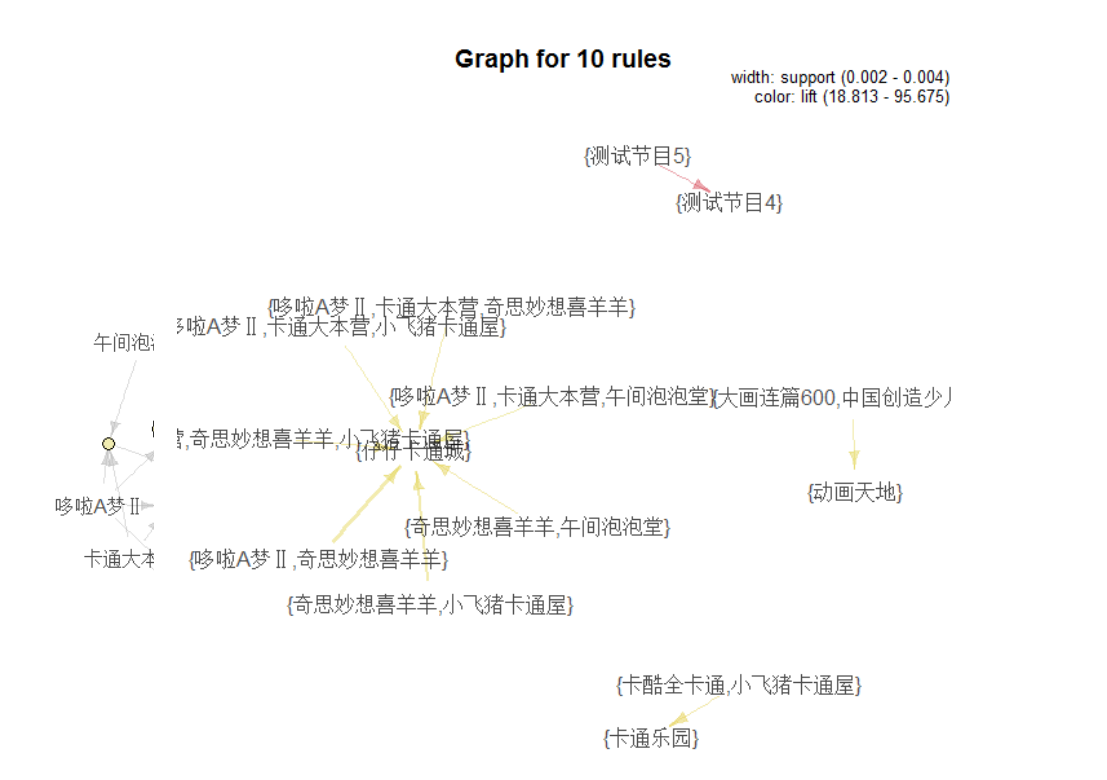


图 33

(1) 对大画连篇 600，中国创造少儿版，动画天地进一步分析

节目	频道	主要播出时间	平均时长(s)	总时长(s)	收看频次(次)
大画连篇 600	金鹰卡通	18 : 00-19:00	925	2744592	2968
中国创造少儿版	卡酷卡通	18 : 30-19:30	589	1438934	2442
动画天地	优漫卡通卫视	18 : 00-19:30	597	2513023	4210

表 11

(表格信息来自网页查询和已有数据集)

由表中可以看出这三档节目均为少儿卡通节目，播出时间、受众群体相似，所以看出喜欢大画连篇 600 的一般也会喜欢中国创造少儿版、动画天地，三者之间存在密切的关联关系。

另外，三者中动画天地的收视总时长最长，且收看频次最高，但因为其播出时长要多于其他两个节目，所以还不能判定是否为最热门节目，需要进一步分析。

(2) 对铠甲勇士、积木世界的童话故事、小肥猪卡通屋、仔仔卡通城、奇思妙想喜洋洋、哆啦 A 梦、卡通大本营分析

节目	频道	主要播出时间	平均时长(s)	总时长(s)	收看频次(次)
奇思妙想喜洋洋	卡酷卡通	7:00-11:00	856.4	4234784	4945
哆啦 A 梦 II	卡酷卡通	11:00-14:00	608	3878206	6379
铠甲勇士	卡酷卡通	14:00-15:00	646.1	2322243	3594
积木世界的童话故事	卡酷卡通	15:00-16:30	622.9	1698525	2727
仔仔卡通城	金鹰卡通	9:00-12:00	870.8	4623072	5309
午间泡泡堂	金鹰卡通	12:00-13:00	919.1	2317867	2526



小飞猪 卡通屋	金鹰卡通	13:00-17:00	710.7	5391884	7587
卡通大本营	优漫卡通 卫视	9:00-13:00	767.6	7397638	9637

表 12

（表格信息来自网页查询和已有数据集）

由此我们可以看出，铠甲勇士、积木世界的童话故事、小肥猪卡通屋、仔仔卡通城、奇思妙想喜洋洋、哆啦 A 梦、卡通大本营这几台节目之间存在着密切的联系。比如喜欢看哆啦 A 梦、卡通大本营和奇思妙想喜羊羊的一般也会喜欢看仔仔卡通城；喜欢看卡通大本营和奇思妙想喜羊羊的一般也看仔仔卡通城；喜欢看卡通大本营和奇思妙想喜羊羊和小飞猪卡通屋的一般也看仔仔卡通城；喜欢看午间泡泡堂、奇思妙想喜羊羊的人一般也喜欢看仔仔卡通城；喜欢奇思妙想喜羊羊的一般会看哆啦 A 梦；看积木世界童话故事和仔仔卡通城的也看小飞猪卡通屋等等。

这是由于这些节目类型、性质相似，主要分布在三个收视频道，播出时间交叉重叠造成的，观众往往同时喜欢其中的多个节目，他们会在一台节目的广告间隙进行换台收看另外一档几乎同时播出的节目，或者在一档喜欢的节目结束时候接着收看另一档节目。

#### 1.3.4. 连续剧类

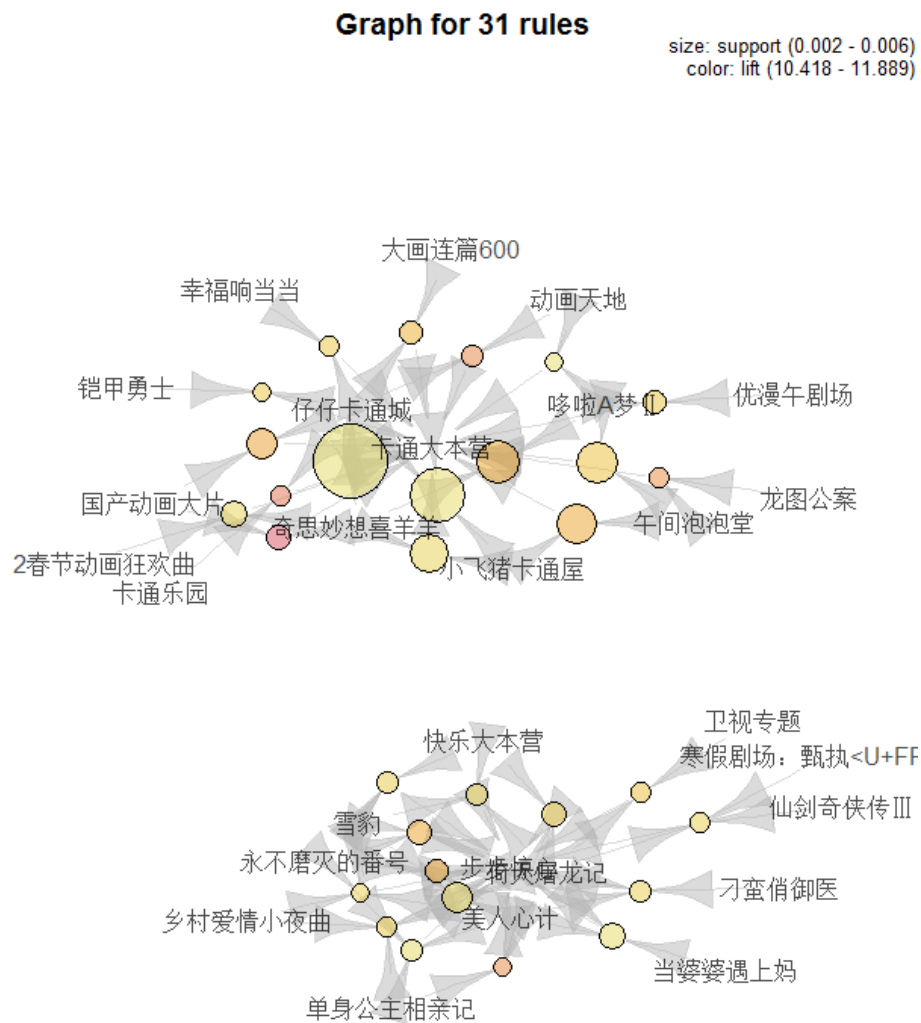


图 34

取提升度排名在 50-80 的规则进行分析（注：前 50 个规则中没有关于连续剧类的关联规则）

我们只要分析图右下角部分。

分析美人心计、快乐大本营、倚天屠龙记、步步惊心、永不磨灭的番号的数据，发现他们的播出时间大致均为下午 13:00-17:00，且同时在多个频道滚动播出；喜欢其中一个节目的观众通常也可能会喜欢另外四个中的节目。

## 1、 用户行为序列图

我们基于每个用户的行为制作了行为序列图，截图中每一行为一个用户，蓝色代表换台，红色代表关机。

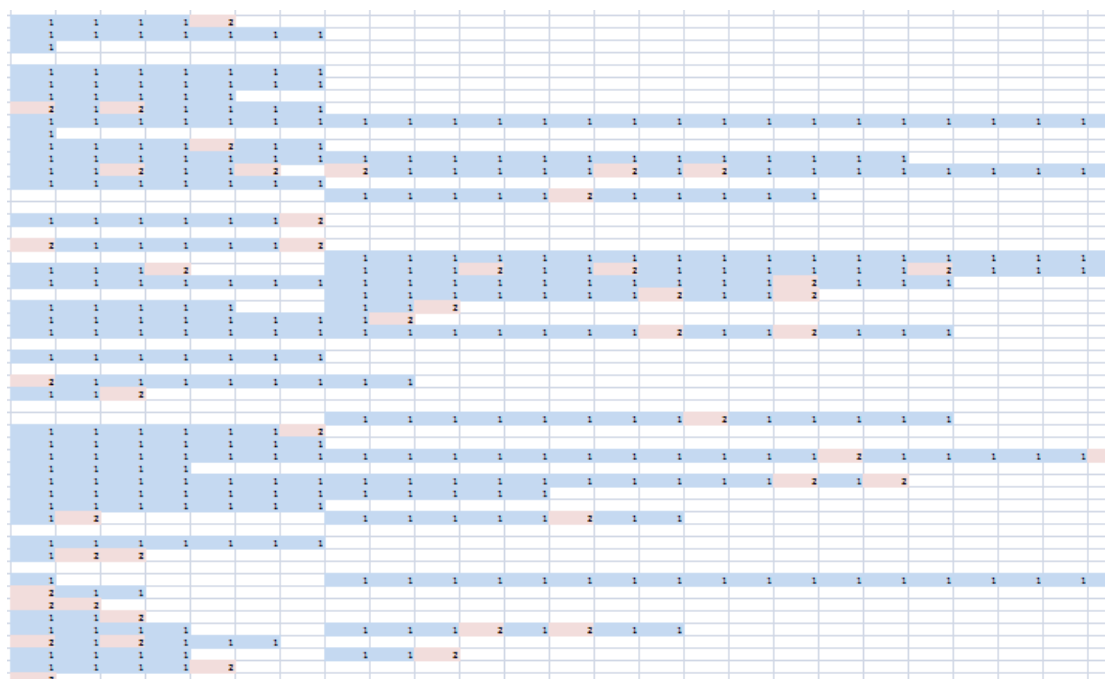
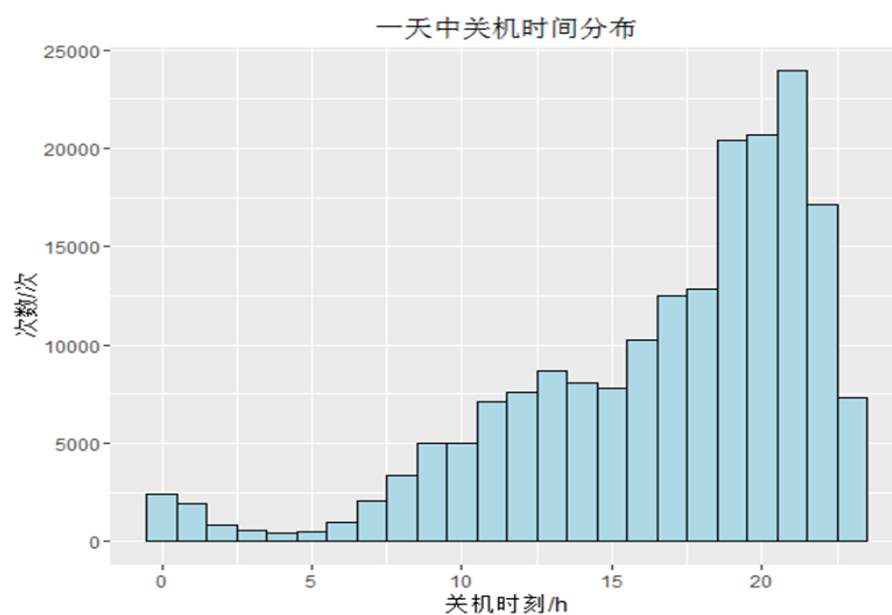


图 35

## 2. Session 分析

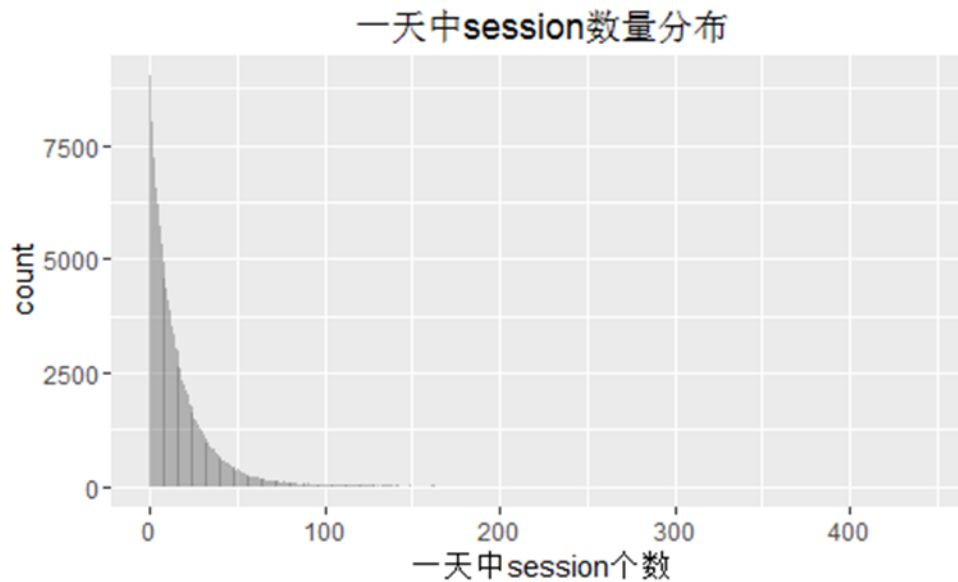
### 2.1 一天中 session 分布时间



图表 36

从图中可以看到，一天中关机次数最多的时刻集中在 18 :00-21 :00，关机次数最低的时刻分布在 0 : 00-7 : 00

## 2.2 . 一天中 session 个数分布



图表 37

一天中关机次数集中分布在 0-50 次之间，其中一天关机次数在 0-10 次的用户最多。

而节目出现 session 个数排名前 20 为：

Program	session_count	Program	session_count
星光大道-2011 年度总决赛	39387	新闻直播间	15990
乡村爱情小夜曲	37754	刁蛮俏御医	15028
幸福三颗星	21296	新闻联播	14846
东方购物	20696	宫锁珠帘	14805
好享购	20120	爱情睡醒了	14539
中国男子篮球职业联赛第 28 轮	19710	美人心计	13826
中国新闻	18816	幸福剧场：怪侠欧阳德	13778
2012 年春节联欢晚会	18769	仙剑奇侠传Ⅲ	13538
城市第一剧场：誓言今生	17314	西游记	13183
新闻直播间	15990	悬崖	13098

table 1 1

频道出现 session 的个数排名前 20：

channel	session_count	channel	session_count
CCTV-3	97700	CCTV-新闻	43575
CCTV-1	73185	江苏卫视	42763
CCTV-4	72156	卡酷卡通	40925
CCTV-6	69363	南京新闻	40915
CCTV-5	60115	CCTV-7	40378
江苏城市	59810	优漫卡通卫视	39807
湖南卫视	51214	江苏影视	38856
CCTV-2	49335	安徽卫视	33220
CCTV-8	46507	北京卫视	32905
		CCTV-10	32556

table 1 2

## 2.3 . 电视节目名称可视化

1. 使用 JiebaR 完成对电视节目名称的分词，去掉没用的停用词。
- 2 . 用 table 统计每个词出现的频率，排序。
- 3 . 导入 Tagxedo 制作云图。



图表 38

可以看到节目名称中出现“场”“情”“家”“新闻”“曲”“品”等非常多，节目类型主要为新闻、综艺、连续剧等；且大概是由于在春节附近，大量的“春”“乐”“年”“节”以及感情色彩比较积极的词语。

### 3 收视率分析

#### 3.1 定义

$$\text{收视率} = \frac{\text{指定时段内收看某一节目的观众收看总时长}}{\text{该时间段所有观众人数} \times \text{该时段时长}}$$

**热门节目：**所有节目中，收视率排名在前 20% 的节目

**优势节目：**同一个频道下，收视率排名在前 20% 的节目

#### 3.2. 节目收视率

##### 3.2.1 整体分析

选择 15min 为时间间隔比较合适，将一天 24 小时划分为 96 个时段。统计在每个时段内有收视记录的所有节目，分别计算出它们的收视率，得到一天内所有节目的收视率变化（7:00~9:00 如下表）。

7:00 到 7:15	朝闻天下	螳螂	星光大道-2011 年度总决赛	早安江苏	中国新闻
	0.025271	0.017447	0.016911	0.016686502	0.013748355
7:15 到 7:30	朝闻天下	早安江苏	螳螂	星光大道-2011 年度 总决赛	亲情树
	0.032273	0.021635	0.016265	0.015900379	0.014152483
7:30 到 7:45	朝闻天下	早安江苏	第一时间	2012 年春节戏曲晚会	螳螂
	0.037066	0.025534	0.015871	0.015867633	0.01468496
7:45 到 8:00	朝闻天下	早安江苏	2012 年春节戏曲 晚会	第一时间	亲情树
	0.041654	0.024853	0.017081	0.016564136	0.01502614

8:00 到 8:15	朝闻天下	2012 年春节戏曲 晚会	早安江苏	中国新闻	第一时间
	0.045989	0.020062	0.019829	0.019499084	0.018734246
8:15 到 8:30	朝闻天下	中国新闻	第一时间	早安江苏	2012 年春节戏曲 晚会
	0.044047	0.022532	0.020242	0.016460029	0.016415396
8:30 到 8:45	朝闻天下	中国新闻	第一时间	甲方乙方	2012 年新春歌会
	0.038143	0.025119	0.020206	0.016831433	0.013793027
8:45 到 9:00	朝闻天下	中国新闻	2012 年新春歌会	甲方乙方	第一时间
	0.029819	0.017811	0.015991	0.01570044	0.014770983

table 1 3

该时间段收视率最高的是朝闻天下，且收视率稳定。其次是江苏卫视的一些早间栏目以及中国新闻。

从全部数据来看，在不同时间段收视率排名最高的节目有明显区别，但星光大道和央视春晚以及螳螂占据了大部分时间的收视率第一。其中螳螂是 2011 年一个谍战类电视剧，可见其热度是非常高的

而不同时间段出现的节目个数也有明显差别，画出个数随时段的变化图线如下：

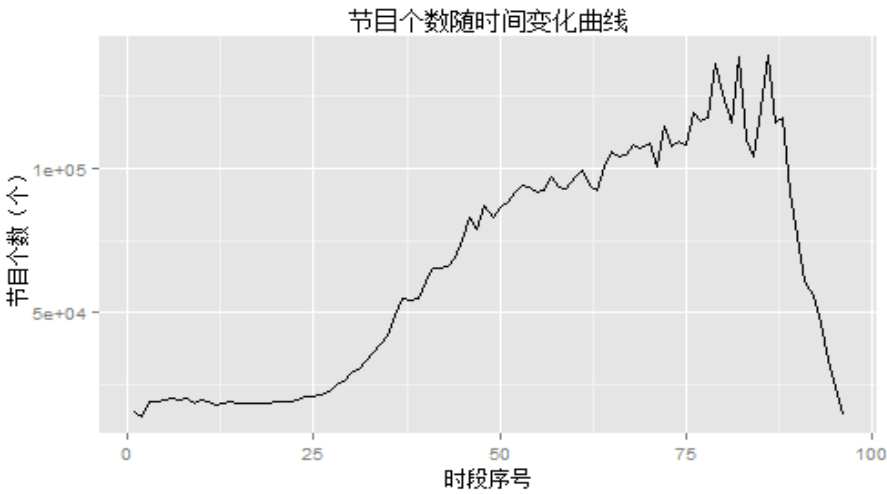


图 39

可见在 18:00 到 22:00 之间电视节目数最多，最多记录到 139390

个节目，而在 00:15 到 00:30 之间记录到节目最少，为 13897 个。

3.2.2 对节目收视率的细化分析

设定一段时间为指定播放节目的完整观看时间，得到所有节目收视率的排名，列出排名前 15 节目的收视率及播出分段。我们发现排名靠前的主要在夜间（中国男子篮球职业联赛第 28 轮、午夜剧场、东方夜新闻等）播放，并且主要播出时段为 1。而总时长排名最为靠前的星光大道、乡村爱情小夜曲等节目虽靠前，但也排到了 40 余名。

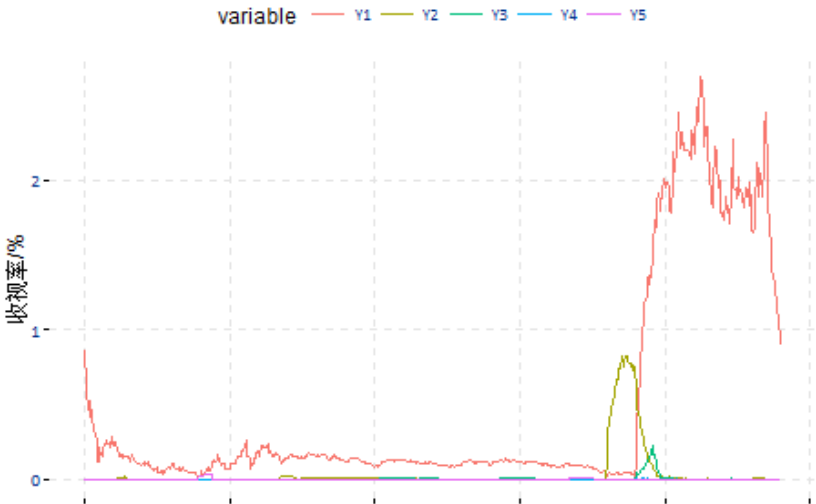
星光大道-2011 年度总决赛	0.007621975	3	44
乡村爱情小夜曲	0.007577391	3	45

table 1 4

我们认为这是由于夜间收看观众的总人数较小，这些节目所占据的时长量相对较大所造成的。

另外我们发现一天中播放的主时段在 3 个以内时，其节目的收视率排名在各个阶段都有分布，但当主时段达到 4 个及其以上时，收视率则主要集中分布于 1000 名以后。考虑有以下两个原因：1．这类节目播放时长短，聚集的人数少，却因为播放时段较多分母很大；2 这类节目本身市场价值就不高，因而被安排在了其他节目的空隙中。

图表 40  
一天中节目收视率变化图（按排名划分）

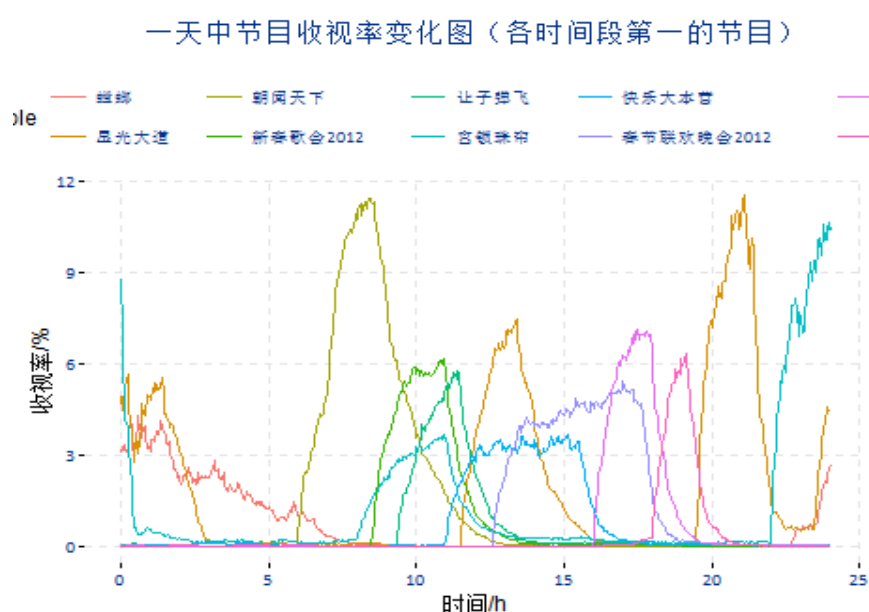




我们先做出收视率排名约间隔 500 名的 5 波节目的收视率变化图(图 40)，其中 Y1-Y5 分别表示前 500 名抽取节目-后 500 名抽取节目，发现无论节目的排名如何，其分布总是大致相似的。类似于时长、频数的收看规律，晚高峰在收视率同样适用。

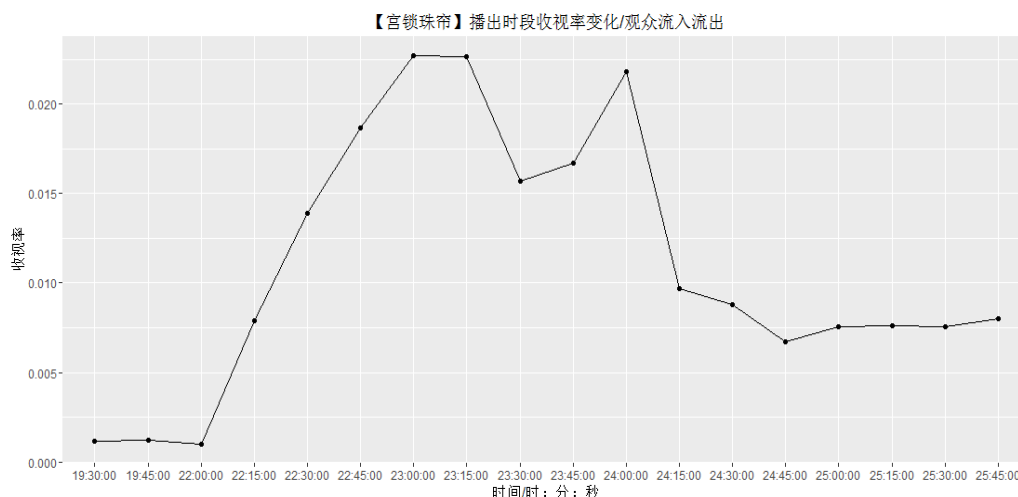
我们再结合 96 个时间段下排名第一次数最多的 10 个节目，画出收视率变化图（图 41）。可以看到：

- 1) 中午及晚上的收视率峰值高于其他时间段；
- 2) 收视率较高的节目一般具有更为尖锐的顶峰( 朝闻天下 ),而较低的节目顶峰则略微平缓( 快乐大本营 ),说明节目的收视率不能仅凭最大值判断，而应该观察一个累积的过程。
- 3) 观察各个峰图，发现上升过程较为迅速，而下降时大多有所拖延，猜测专门等待观看这些高收视率节目的观众较多，并且观看结束时仍存在延迟现象。



图表 41

### 3.2.3 热播电视剧的简单分析



图表 42

查询资料得知,【宫锁珠帘】为 2012 年 2 月份首播的热播电视剧,播出以来收视率持续占据高位。

我们研究该剧播出的收视情况,虽然开始播出时间为 22 时,但开始前已经有不少追剧观众直接在湖南卫视开机等待节目播出。

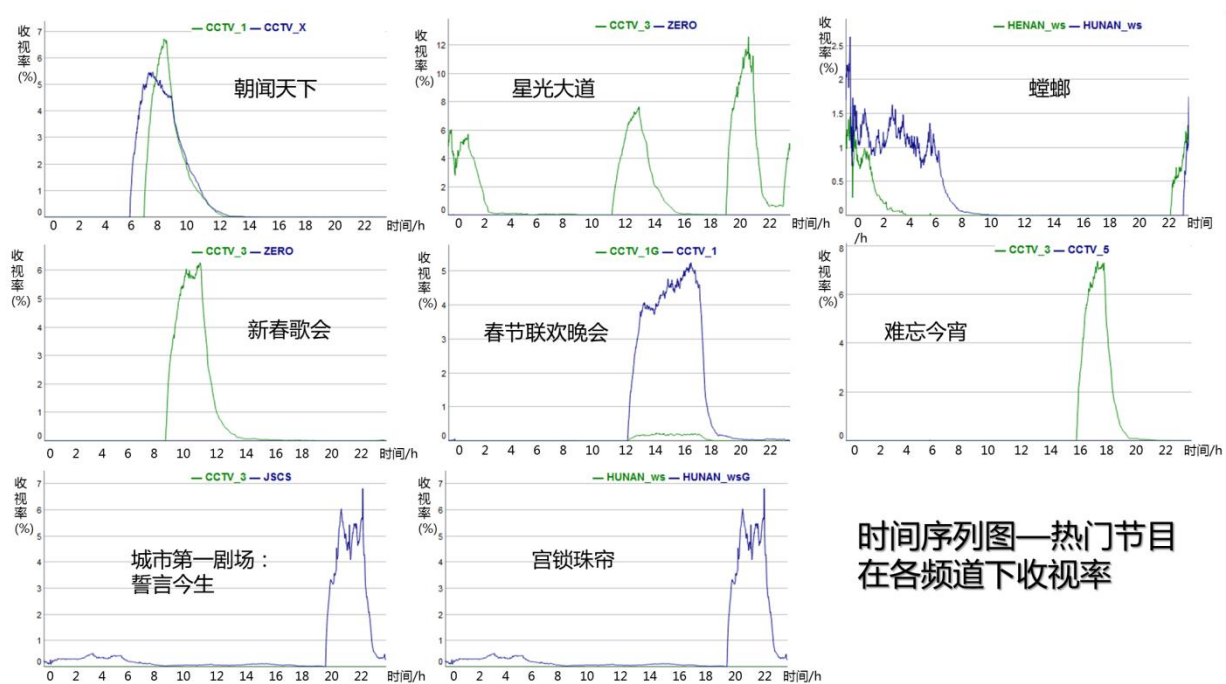
开始播出后大量用户持续涌入,保持稳定的增长,而且大部分都是直接开机,其他来源有【幸福三颗星】【城市第一剧场-誓言今生】等。

23:30 收视率有所下降,这时流失的流向主要为脱口秀节目【非常了得】(另外一档热播节目),很快【非常了得】结束,【宫锁珠帘】又回到收视高峰。

24:00 节目结束,收视率不是瞬间将为 0,而以一定速率下降到一个稳定的水平,深夜仍旧有不少观众回看录播呀。

### 3.2.4 节目-频道时间序列图

以 15 分钟为间隔,我们提取出每个时段收视率第一的节目,画出它们在各个频道收视率的时间序列图,发现:



图表 43

- 1、 几乎每个节目都有零号频道的参与，但其收视率在时间序列图上几乎不可见，说明该频道看似良好的收视情况仅仅是个数的叠加；
- 2、 这些热门节目能检测到的节目基本上与各个时段收视率最高的
- 3、 频道一致：CCTV-1、CCTV-3、湖南卫视、江苏城市
- 4、 时间序列图较为一致的频道实际上是同一类（如：高清与非高清—春节联欢晚会），而分布差异较大者则来自不同频道（如：湖南卫视与河南卫视—螳螂）

### 3.2.4 频道对节目收视率的提升作用【利用假设检验】

电视节目在选择播出平台时，需要衡量不同渠道的市场价值以做出判断，因此判断一个频道对节目收视率的提升作用是很有意义的。接下来我们就将通过假设检验来判断频道对于节目收视率是否存在

影响，是积极影响还是消极影响，从而了解观众对频道的偏好。

以江苏卫视为例，该频道所播出的非诚勿扰节目，会同时在多个频道播出,如下表所示：

时间段	频道及收视率	频道及收视率	频道及收视率	频道及收视率
1	CCTV-1	江苏卫视	江苏卫视高清	
	0.000226879	9.72E-05	8.73E-06	
30	江苏卫视	江苏卫视高清		
	0.000233654	0.000116827		
31	江苏卫视	江苏卫视高清		
	0.000222683	6.52E-05		
50	3D 频道	江苏城市	江苏卫视	
	0.000229517	5.06E-05	6.43E-05	
52	CCTV-1	江苏卫视	南京十八	
	2.69E-05	0.000113557	2.69E-05	
53	CCTV-1	江苏卫视	零频道宣传	南京十八
	2.76E-05	0.000131912	1.11E-05	2.76E-05
54	3D 频道	江苏城市	江苏卫视	
	0.000103308	0.013515313	8.79E-05	
57	江苏卫视	南京教科		
	0.000209904	2.01E-05		
58	江苏卫视	零频道宣传		
	0.000138243	3.50E-05		
85	CCTV-1	江苏卫视	江西卫视	南京十八
	5.01E-05	6.49E-05	5.53E-07	3.57E-05
90	CCTV-1	安徽卫视	江苏卫视	
	6.98E-05	1.71E-05	5.80E-05	

表 44

而江苏卫视播出的晚间新闻也会出现在多个频道播出的情况：

时间段	频道及时长	频道及时长	频道及时长	频道及时长
34	3D 频道	江苏卫视		
	0.00041913	0.004588225		
35	3D 频道	江苏卫视		
	0.0003383	0.003285083		
36	3D 频道	江苏卫视		
	0.00023371	0.001602264		

63	3D 频道	江苏城市	江苏国际	江苏卫视
	0.00056969	0.00025262	0.000374658	0.000163626
64	CCTV-1	江苏卫视		
	3.62E-05	0.000164978		
88	CCTV-1	江苏卫视	南京十八	
	1.86E-05	5.58E-05	1.24E-05	
96	CCTV-1	CCTV-1 高清	江苏卫视	
	0.00020954	8.48E-05	0.001813916	

那么，对于江苏卫视，我们可以考察该频道在一天内所有时段播出的节目，如果第  $i$  个时段江苏卫视播出的节目  $\text{pro}$  非独播，我们就计算该时段内， $\text{pro}$  在所有频道的平均收视率  $\bar{r}_i$ ，以及该节目在江苏卫视的收视率  $r_i$ ，则我们可以比较一天内所有的  $r_i$  及  $\bar{r}_i, 1 \leq i \leq 96$ ,

记  $r_i > \bar{r}_i$  为正， $r_i < \bar{r}_i$  为负，将此符号结果即为  $l_i$ ，则可对一系列的  $l_i$  进行符号检验。

因为实际数据中包含了不止一天的数据，所以同一频道同一个时段内可能出现多个非独播节目，故而有些频道的检验总数可达到两百多个。

江苏卫视的非独播时段及检验量如下：

时段	节目平均收视率	该频道收视率	检验量
1	0.000110926	9.72E-05	0
30	0.000175241	0.000233654	1
31	0.000143949	0.000222683	1
48	0.000106087	0.000132218	1
50	0.000114796	6.43E-05	0
52	5.58E-05	0.000113557	1
53	4.96E-05	0.000131912	1
54	0.004568835	8.79E-05	0
56	0.00326856	0.000115145	0
57	0.000115006	0.000209904	1
58	8.66E-05	0.000138243	1
59	0.000575819	0.000148065	0

60	0.000463831	0.000145817	0
61	0.000416796	0.000145726	0
62	7.93E-05	0.000156206	1
63	8.97E-05	0.000138537	1
65	0.000193503	9.32E-05	0
66	0.000165996	9.93E-05	0
67	0.000170329	9.90E-05	0
68	0.000156249	0.000115521	0
69	8.75E-05	4.43E-05	0
70	4.55E-05	8.26E-05	1
72	5.31E-05	9.20E-05	1
73	0.00015447	0.000160393	1
77	0.000104519	0.000164828	1
78	0.000104751	0.000154107	1
79	0.000226179	0.000445341	1
80	9.94E-05	0.000156594	1
81	7.25E-05	0.000130274	1
83	5.65E-05	7.52E-05	1
85	3.78E-05	6.49E-05	1
86	2.76E-05	4.31E-05	1
87	3.16E-05	5.45E-05	1
90	4.83E-05	5.80E-05	1

以此类推，检验步骤如下：

定义 $r_{pro}$ 为节目 $pro$ 在某一时段所有频道的收视率， $r_{chpro}$ 为节目 $pro$ 在频道 $ch$ 同样时段播出时的平均收视率，接下来对比 $r_{pro}$ 与 $r_{chpro}$ 的大小，记 $p$ 为 $r_{pro} > r_{chpro}$ 的概率。

$$H_0: p = 0.5 \quad H_1: p > 0.5$$

接下来找到 $ch$ 在所有时间段出现的所有节目，记出现的不同时段的节目总数为 $n$ 。对比每个节目的 $r_{chpro}$ 与 $r_{pro}$ ，记 $r_{pro} > r_{chpro}$ 的节目个数为 $n_0$ 。

$$\text{则检验 } p\text{-value} = \sum_{i=n_0}^n \binom{n}{i} (0.5)^i。$$

当  $n > 100$ , 选择显著性水平为 0.001, 当  $n < 100$ , 选显著性水平为 0.05。

如果  $p$  值小于 0.05 则拒绝  $H_0$ , 说明频道对节目收视率有明显提升作用, 否则不拒绝零假设, 即不能认为频道对节目收视率有明显提升作用。

name	rate	备 择选项	检验 p 值	r	success	estimated
CCTV -3	0.04 3898	greater	3.70E- 16	1 71	137	0.8011695 91
CCTV -1	0.03 2901	greater	3.78E- 15	6 52	425	0.6518404 91
CCTV -6	0.03 2892	greater	0.4627	1 14	58	0.5087719 3
CCTV -4	0.03 0747	greater	8.71E- 05	2 75	169	0.6145454 55
湖南 卫视	0.02 4774	greater	< 2.2e-16	3 40	314	0.9235294 12
CCTV -5	0.02 3852	greater	5.00E- 01	3 7	19	0.5135135 14
CCTV -新闻	0.02 2017	greater	0.0199 2	2 28	130	0.5701754 39
CCTV -2	0.02 1294	greater	0.0007 252	4 1	31	0.7560975 61
CCTV -8	0.02 0724	greater	8.13E- 08	1 78	124	0.6966292 13
江苏 卫视	0.01 5287	greater	< 2.2e-16	4 40	345	0.7840909 09
安徽 卫视	0.01 4032	greater	0.0258 4	3 24	180	0.5555555 56
CCTV -7	0.01 34	greater	1.16E- 06	1 65	113	0.6848484 85
浙江 卫视	0.01 2304	greater	0.0012 71	3 58	208	0.5810056
北京 卫视	0.01 1975	greater	< 2.2e-16	3 60	270	0.75
广西 卫视	0.01 1449	less	< 2.2e-16	1 43	41	0.2867133
东方 卫视	0.01 117	greater	< 2.2e-16	3 80	296	0.7078947
CCTV	0.01	greater	1.49E-	3	193	0.6348684

-少儿	0372	ter	06	04		
CCTV	0.01	grea	0.0001	2	108	0.6305419
-10	036	ter	219	03		
江西	0.01	grea	0.0067	2	125	0.5868545
卫视	0185	ter	25	13		
山东	0.00	grea	0.0093	1	75	0.6097561
卫视	9872	ter	47	23		
天津	0.00	grea	0.0003	2	129	0.6201923
卫视	9699	ter	222	08		
湖北	0.00	grea	1.42E-	2	149	0.6563877
卫视	9553	ter	06	27		
辽宁	0.00	less	3.03E-	1	38	0.2196532
卫视	9508		14	73		
CCTV	0.00	grea	0.0002	2	24	0.8275862
-11	9352	ter	731	9		
黑龙	0.00	grea	2.19E-	2	169	0.6869919
江卫视	9121	ter	09	46		
四川	0.00	less	0.2744	1	5	0.4545455
卫视	8981			1		
东南	0.00	less	0.0001	5	14	0.2545455
卫视	7991		776	5		
重庆	0.00	grea	0.0073	2	19	0.76
卫视	7724	ter	17	5		
贵州	0.00	less	0.0669	2	7	0.3181818
卫视	728			2		
云南	0.00	less	0.0319	7	29	0.3866667
卫视	7111		7	5		
旅游	0.00	less	0.0001	7	21	0.2876712
卫视	6905		857	3		
河北	0.00	less	0.3776	4	19	0.4634146
卫视	6902			1		
吉林	0.00	grea	0.6873	6	32	0.4776119
卫视	6874	ter		7		
青海	0.00	grea	0.0058	7	50	0.6493506
卫视	6453	ter	36	7		
广东	0.00	grea	<2.2e-	2	190	0.8837209
卫视	6292	ter	16	15		
河南	0.00	less	<2.2e-	1	24	0.1472393
卫视	5479		16	63		
山西	0.00	grea	0.0073	2	19	0.76
卫视	5312	ter	17	5		
CCTV	0.00	less	0.0680	2	10	0.3448276
-12	2362		2	9		



宁夏卫视	0.002315	greater	1.15E-05	106	75	0.7075472
CCTV-音乐	0.000778	less	0.1239	27	10	0.3703704
兵团卫视	0.000597	less	0.2148	40	17	0.425

表中标红的频道是具有显著提升作用的，标蓝的频道是有显著负面作用的。可以看出提升作用最显著的几个频道如下：

name	rate	备择选项	检验 p 值	n	success	estimated
CCTV-3	0.043898	greater	3.70E-16	171	137	0.801169591
CCTV-1	0.032901	greater	3.78E-15	652	425	0.651840491
湖南卫视	0.024774	greater	< 2.2e-16	340	314	0.923529412
江苏卫视	0.015287	greater	< 2.2e-16	440	345	0.784090909
北京卫视	0.011975	greater	< 2.2e-16	360	270	0.75
东方卫视	0.01117	greater	< 2.2e-16	380	296	0.7078947
广东卫视	0.006292	greater	<2.2e-16	215	190	0.8837209

同时段播出的相同节目，如果在这些频道播出，则其收视率有较大概率会高于平均水平。说明这些频道对于节目会有较强的提升作用。可以算作品牌塑造较好的频道，使得观众有更高的热衷度，也就是说这些频道提供了更具有市场价值的播出平台。

相对应的，几个负面作用比较显著的频道如下：

name	rate	备择选项	检验 p 值	n	success	estimated
广西卫视	0.011449	less	< 2.2e-16	143	41	0.2867133
辽宁卫视	0.009508	less	3.03E-14	173	38	0.2196532
东南卫视	0.007991	less	0.0001776	55	14	0.2545455
云南卫视	0.007111	less	0.03197	75	29	0.3866667
旅游卫视	0.006905	less	0.0001857	73	21	0.2876712
河南卫视	0.005479	less	<2.2e-16	163	24	0.1472393

这些频道在所有的央视频道以及卫视中并不完全属于收视率低的频道。例如广西卫视和辽宁卫视，他们的综合收视率比很多有显著提升作用的频道要高，由此可知，这两个频道播出的节目本身收视率不低，但是相对这些节目播出的其他频道，这两个频道的观众热衷度不高。也就是说，这两个卫视可能是成功地通过节目来拉高了频道的收视水平。而剩下的几个频道本身综合收视率都不高，属于品牌营造不太成功的频道，观众对于频道本身没有热衷度，播出的节目综合收视率也一般。

并且从综合收视率和频道对节目的提升或负面作用可以判断出频道的节目策略：

综合收视率	作用	类型
高	提升作用	频道品牌营造成功，观众认可频道，且播出节目质量高
高	负面作用	观众对频道没有明显热衷，但播出了热门节目，通过节目成功提高收视率
低	提升作用	播出节目非热门但很有频道特色，有小众的人群热衷在这些频道收看某些节目
低	负面作用	频道没有明显的观众热衷度，且没有播出热门节目

### 3.3 频道收视率

#### 3.3.1 地方卫视以及央视求所有时段平均收视率表

排序	频道	收视率	排序	频道	收视率	排序	频道	收视率
----	----	-----	----	----	-----	----	----	-----

频道收视率的定义与之前节目的类似，从得到的频道收视率表格中可以看出央视及各大卫视的收视率在各时段均领先，接下来我们对各地方卫视以及央视求所有时段平均收视率，并得到收视率排名。

1	CCTV-3	0.043897	19	CCTV-10	0.010360	37	河南卫视	0.005479
2	CCTV-1	0.032900	20	江西卫视	0.010184	38	山西卫视	0.005312
3	CCTV-6	0.032892	21	山东卫视	0.009871	39	CCTV-12	0.002361
4	CCTV-4	0.030746	22	天津卫视	0.009699	40	CCTV-9 纪录	0.002322
5	湖南卫视	0.02477	23	湖北卫视	0.009553	41	宁夏卫视	0.002315
6	CCTV-5	0.023851	24	辽宁卫视	0.009507	42	湖南卫视高清	0.001766
7	CCTV-新闻	0.022017	25	CCTV-11	0.009351	43	深圳卫视高清	0.001623
8	CCTV-2	0.021294	26	黑龙江卫视	0.009120	44	东方卫视高清	0.001575
9	CCTV-8	0.020724	27	四川卫视	0.008981	45	CCTV-1 高清	0.001533
10	江苏卫视	0.015286	28	东南卫视	0.00799	46	北京卫视高清	0.001481
11	安徽卫视	0.014032	29	重庆卫视	0.007724	47	黑龙江卫视高清	0.001453
12	CCTV-7	0.013400	30	贵州卫视	0.007279	48	江苏卫视高清	0.001370
13	浙江卫视	0.012303	31	云南卫视	0.007110	49	浙江卫视高清	0.001318
14	北京卫视	0.011975	32	旅游卫视	0.006905	50	广东卫视高清	0.001231
15	优漫卡通卫视	0.011953	33	河北卫视	0.006902	51	CCTV-音乐	0.00077
16	广西卫视	0.011448	34	吉林卫视	0.006873	52	兵团卫视	0.000596
17	东方卫视	0.011170	35	青海卫视	0.006452	53	新疆卫视	0.000313
18	CCTV-少儿	0.010371	36	广东卫视	0.00629	54	甘肃卫视	0.000281

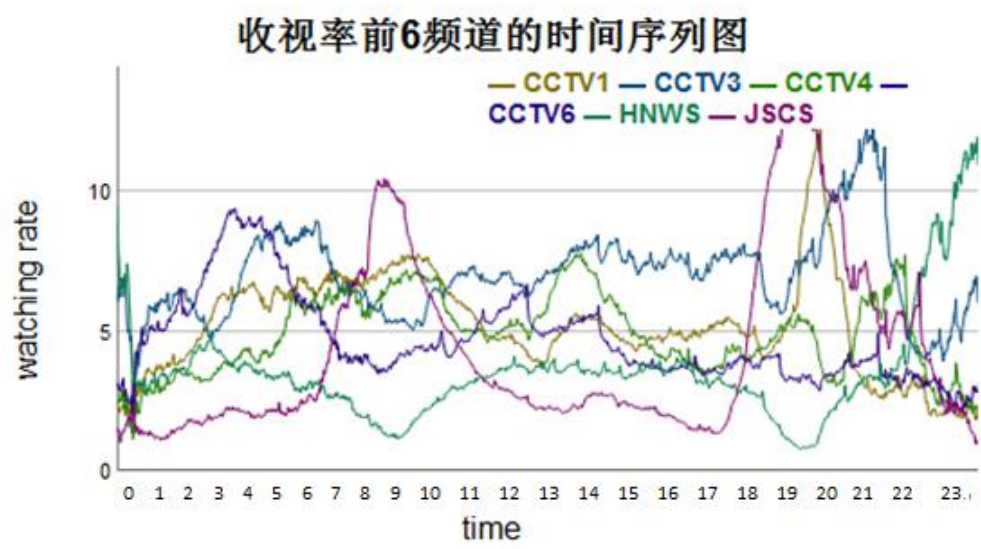
表格 15

### 3.3.2 基于频道的流入流出

综合全天各个时段，收视率最高的 6 个频道分别为：CCTV-1、CCTV-3、CCTV-4、CCTV-6、湖南卫视及江苏城市，它们全天的收视率变化情况如下图。

无优势节目播出时，收视率较为平稳，如 15:00-17:00 这一时段，

各频道收视率能基本反应其受欢迎度：此时收视率由频道本身决定。但当某个频道出现优势节目时，该频道会有大量观众流入，其他频道则流出，如 18:00-22：00 一段，竞争非常激烈：起初，江苏城市受其优势节目（零距离）影响收视率大幅增加，与此同时其他各频道收视率降低；随后在 20:00 左右，尽管 CCTV-1 的收视率有所上升，但难以阻止观众对 CCTV-3 优势节目（星光大道）的喜爱，不多时便与其他各节目降至一个低点。



图表 44

为了进一步确认收视率的明显变化（观众的流入流出）由优势节目的播出造成，我们提取每个热门频道前 20% 的节目做具体分析：

频道	CCTV-1	CCTV-3	CCTV-4	CCTV-6	湖南卫视	江苏城市
前 20% 节目数	8	8	7	10	6	5
节目收视率比	79.1%	84.3%	67.4%	78.0%	93.5%	64.9%

表格 16

可以看出最热门的 6 个频道中大约 80% 的收视率是由收视率排名前 20% 的节目贡献的，与前文提到的二八原则是一致的。我们继

续观察每个频道收视率前 20%的节目的具体情况：

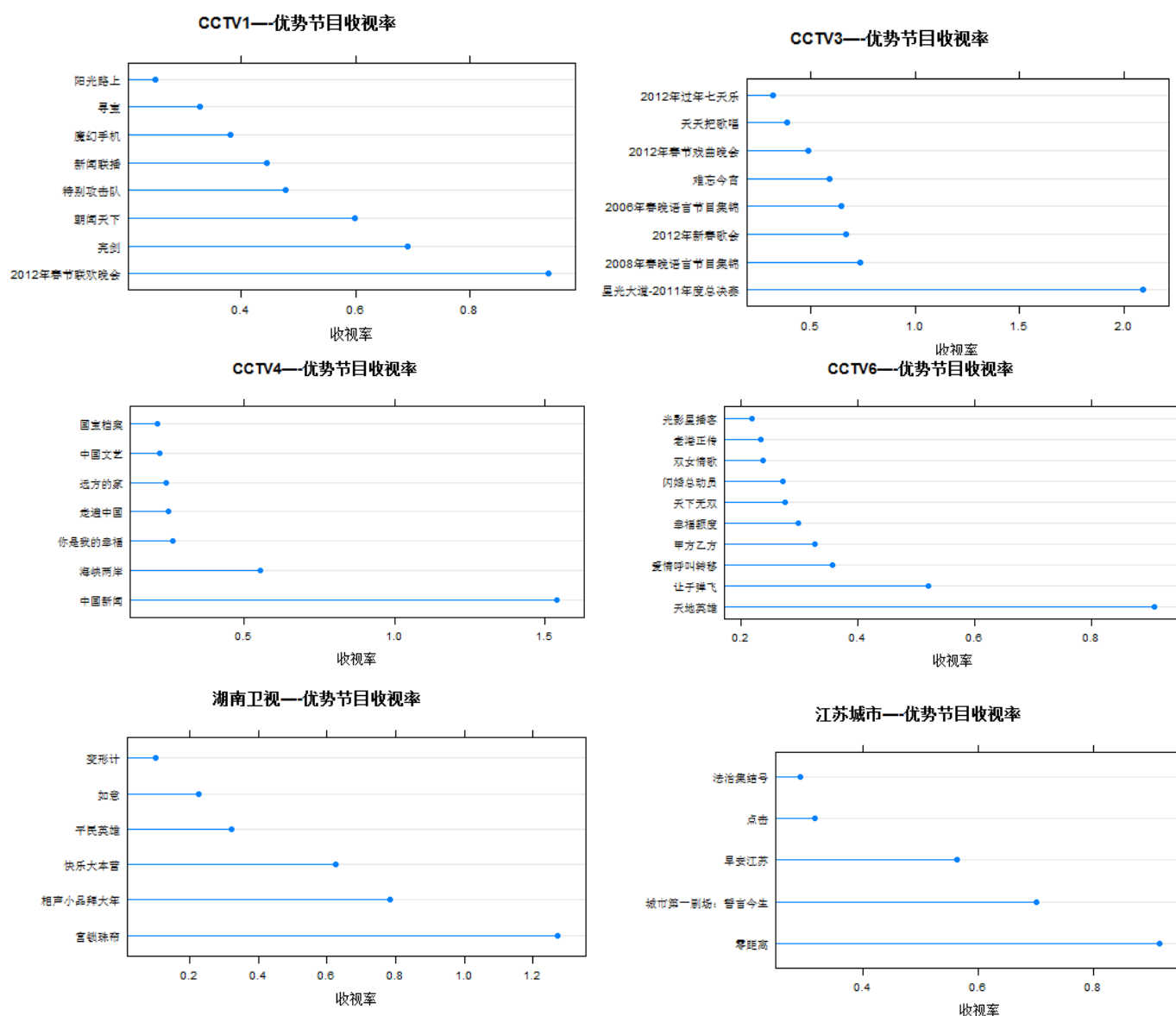


图 45

发现各频道的优势节目与前文的热门节目基本保持一致。结合热门节目的收视率时间序列图，证实了造成观众明显流入流出的原因及热门节目的出现：热门节目频道的收视率图像在波动处与该节目的收视率图像基本保持一致。

### 3.3.3 频道收视率模型

1) 回归模型

目的：

通过建立某频道收视率关于该频道下各节目收视率的模型 ,我们希望找到频道与该节目收视率的数值关系 ,并在没有所有节目收视率的情况下仍然能模拟出频道的收视率。

建立模型 ( CCTV-3 ):

i、与优势节目收视率

利用前文验证过的二八定律 ,我们选择该频道收视率排名前 20% 的节目 ( 贡献率达 84.3% ) 作为解释变量建立回归模型 1

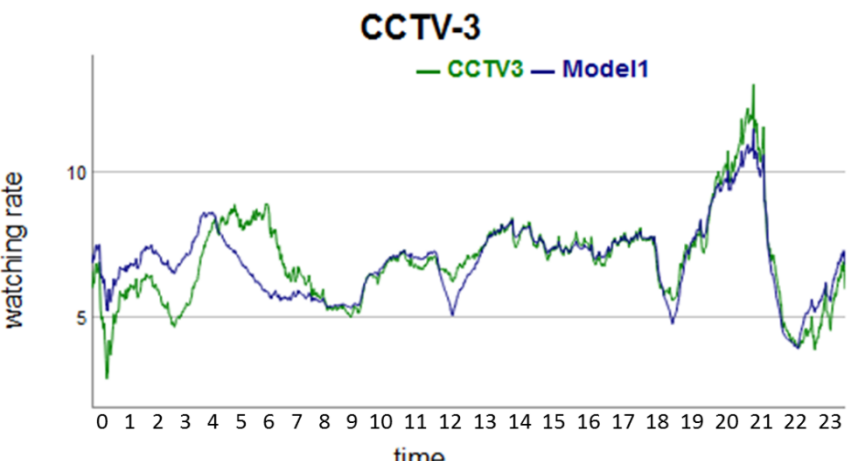


图 48

( AIC=3614.9 , 1431 的自由度下残差值为 1023.5 )。观察模型 1 与原始频道数据的收视率时间序列图 , 我们

发现 8 点后模型的拟合效果较为良好 , 而凌晨段的差异较大 , 为进一步分析波动的原因 , 我们画出前 20% 节目的收视率累积图。

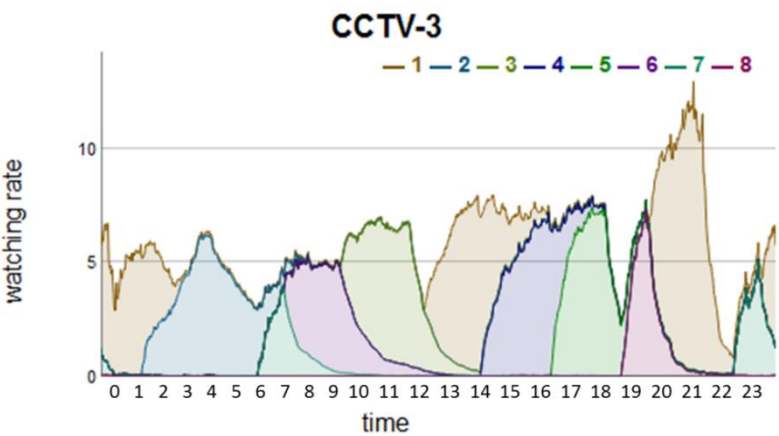


图 49

发现差异较大的时间段主要播放的节目主要是该频道收视率排名 1、2 的节目 , 故而加入

这两个节目各自的二次项与二者间的交互项，重新建立模型 2（ $AIC=2767.2$ ，1428 的自由度下有残差 565.74，比模型 1 有很大改善，但残差序列非正态）。得到的模型与原频道数据对比图及模型表达式如下。

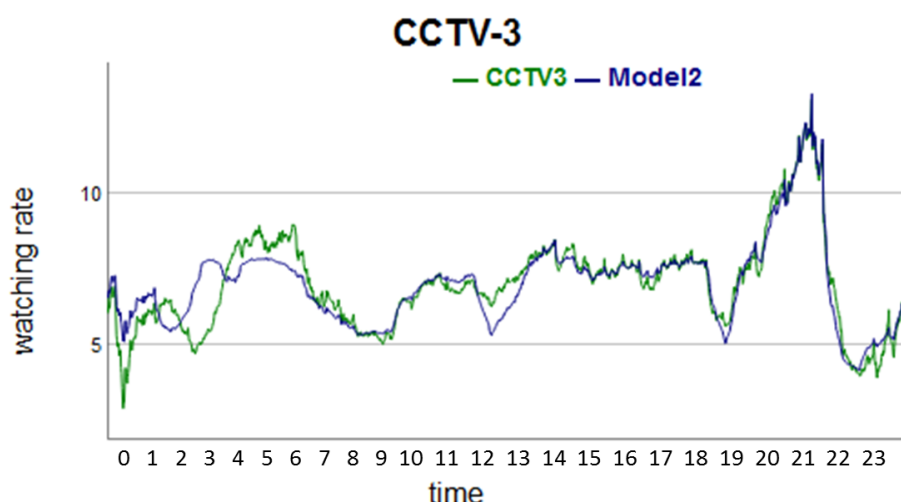


图 50

$$\begin{aligned} \text{model2} = & 1\text{st 收视率} \times 0.357 + 2\text{nd 收视率} \times 1.883 + 3\text{rd 收视率} \times 0.524 \\ & + 4\text{th 收视率} \times 0.528 + 5\text{th 收视率} \times 0.514 + 6\text{th 收视率} \times 0.320 \\ & + 7\text{th 收视率} \times 0.276 + 8\text{th 收视率} \times 0.592 + 1\text{st 二次项} \times 0.030 \\ & - 2\text{nd 二次项} \times 0.21200379 - 1\text{st 与 } 2\text{nd 交互项} \times 0.409 + 3.829 \end{aligned}$$

## ii、与播放时间段为一的节目

由于在上面的模型中，贡献率最大的变量并非收视率最高的节目，而是如第 3-5 这些全天只有一个主要播放时段的节目，说明主播放时段也很大的影响了节目对频道收视率的贡献。

我们推测对频道收视率贡献最大的节目收视率，来自只有一个主播放时段的节目。故而对这些只有一个主播放时段唯一的节目按收视率排名，重新建立模型 3（ $AIC=5052$  残差=2778 自由度=1431），



残差比模型 2 大了 4 倍左右，改进失败。考虑是因为收视率排名第一的星光大道（两个主播放时段），被排除的原因。

**结论：**

- A、 影响频道收视率的主要贡献：节目收视率、节目主播放段；
- B、 收视率越高的节目，对频道收视率越大；
- C、 收视率情况接近时，仅一个主播放时段的节目具有更大的贡献；
- D、 该模型的思路可用于：已知一个频道的部分高收视率节目情况，估计整个频道的收视率；但只适用于高收视率节目均只有一个主播放时段的情况。

### **3.3.3 热门节目可以提升频道收视率**

**目的与解释：**

在电视频道收益模式中，选择合适的节目有助于提升频道的收视率，并吸引更多的广告商，从而获得更多的收益。

根据前文的定义，热门是在所有频道中收视率排名靠前的 20%，优势节目则是相对某个确定频道，可以提升该频道收视率的节目。讨论热门节目是否可以提升频道收视率也即热门节目与优势节目是否基本保持一致。若一致，则说明这些节目具有普遍性，适合各类频道；若不一致，则说明热门节目的高收视率来自个别频道的极度优势，不适合绝大部分频道。

在这部分的讨论中，我们先验证热门节目与优势节目基本保持一致的猜测，并希望最后能得到一个挑选节目的策略。

检验：

原用 lattice 画的图表达含义与后 rCharts 新改图重复，删去

随机从热门节目中抽取 150 个节目，非热门节目中抽取 40 个节目，对每一个节目，分别计算它所在频道的全天平均累计收视率，及没有该节目时的平均累计收视率。上图是其中六个节目的示例：左边为热门节目（前 20%），右边为非热门节目（后 80%）。若蓝点均位于红点左侧，说明该节目提升频道收视率。

考虑到部分节目会在某些频道中闪现，与正常收看常识不符，为减少该情况造成的影响，我们将各个频道的平均收视率差异之和为依据判断是否提升（如上图中宫锁珠帘，随位于左侧的多是蓝点，all-except 的差额总和为正，说明节目得到提升），其中按节目名称排名得到下表（Y：热门；improve>0：提升）。

program	hot	improve	Program	hot	improve2
2012 春节动画狂欢曲	Y	0.547	爱情呼叫转移	Y	0.947
2012 非洲国家杯小组赛:加蓬	Y	-0.009	爱情连连看	Y	0.458
2012 年春节联欢晚会	Y	3.153	白领剧场二集：倾世皇妃	N	0.068
2012 年过年七天乐	Y	0.785	蝙蝠侠前传之侠影之谜-上	Y	0.065
A 计划	Y	0.015	蝙蝠侠前传之侠影之谜-中 2	Y	0.037

A 计划II	Y	-0.018	标点 美食(复)	Y	-0.013
CQTV 早新闻	N	-0.098	标点 新娘(复)	N	-0.140
J 品任我游：冒险王	N	0.039	标点 新娘首播	N	-0.124
阿特的奇幻之旅	Y	-0.010	财商先锋首播	N	-0.137
爱和乐	Y	-0.010	超级变变变	Y	0.079

由前文定义，热门节目是收视率前 20%的节目，故取排名前 480 的节目为热门。利用 Pearson 卡方检验，原假设为：热门节目与是否提升收视率无关，备择假设为热门节目提升节目收视率，得到：X-square=31.221，p-value=  $2.3 \times 10^{-8} < 0.01$ ，因而我们有犯错误率不超过 1%的把握认为热门节目确实可以提高频道收视率。

	热门节	非热门	Sum
提升频道收	70	8	76
未提升	40	42	74
sum	110	50	160

提升度：

i) 提升度与收看时刻：

增大样本量后重新做图：利用检测抽取的节目频道信息，画节目提升度相对于播放时段的柱状图（为方便时刻间对比，提升度=该时段提升度/该时刻频次），发现在前文中得到的 12 点、19 点两个小高峰提升均不太明显，而夜里及下午则有较明显提升度。原因除了竞争小，更容易以热门节目胜出外，考虑是否有节目类型（下午连续剧比较多）

的影响，我们进一步研究节目类型与提升度关系。

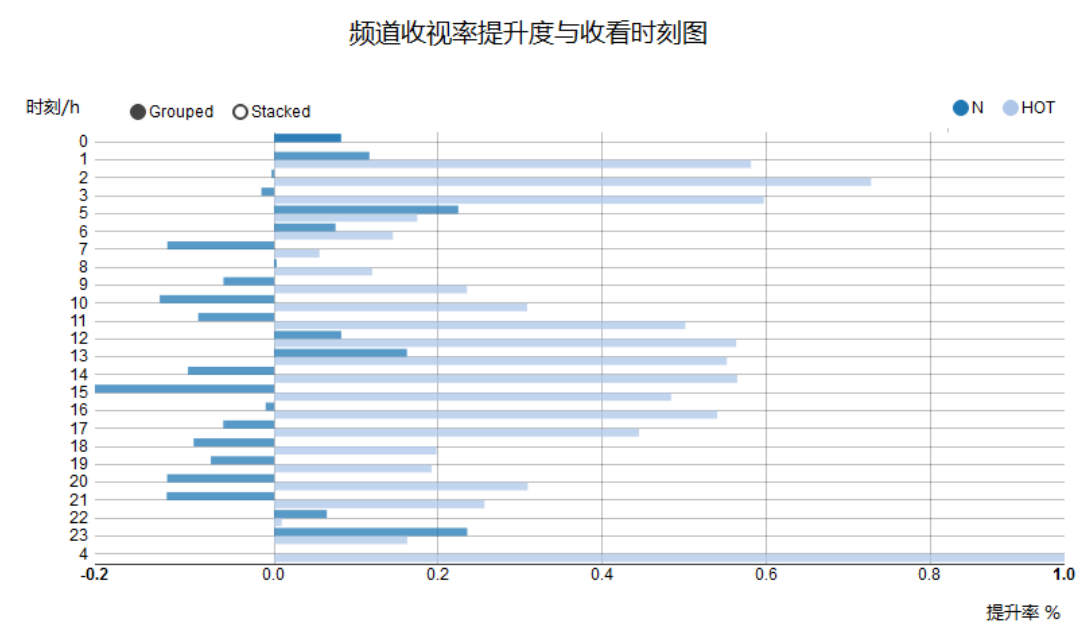
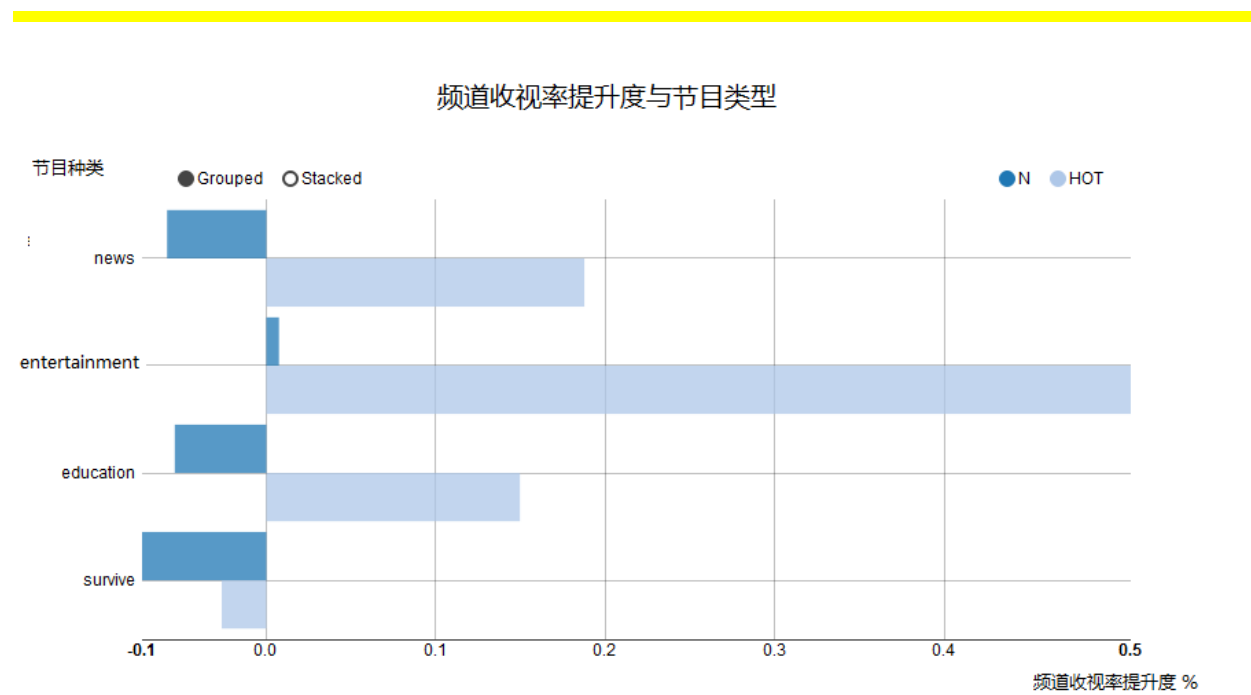


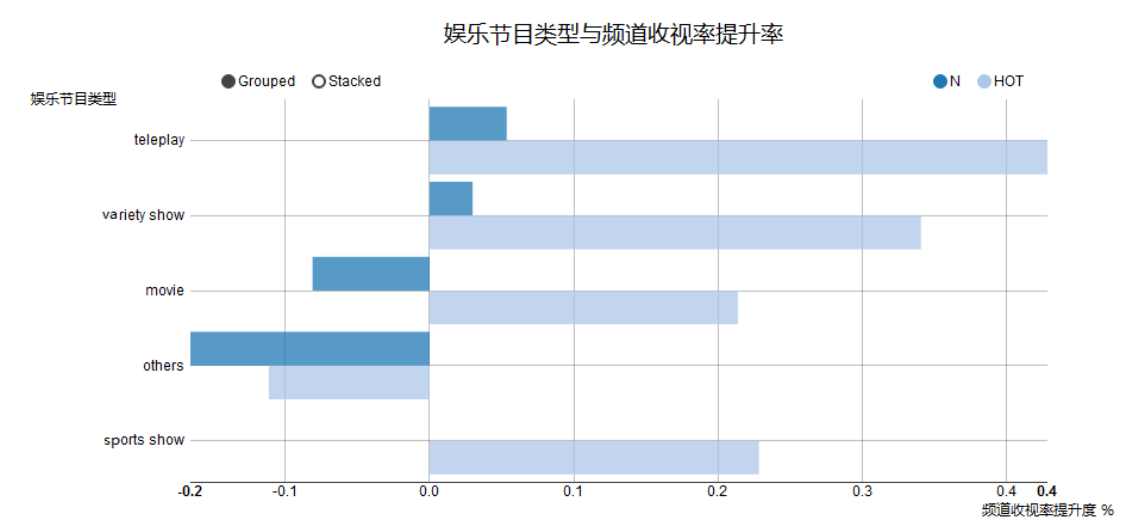
图 52

## ii) 提升度与节目类型

通过节目提升度关于节目类型的柱状图（提升度=该段数下提升度总和/节目类型出现频次），可以看到服务类节目是电视台收入的一项来源，但对频道收视率的负作用比较明显，而提升比较明显的是娱乐类节目（我们组的划分中包含了电视剧），因而后面可以继续考虑给定服务类节目收入、广告收入时，他们与娱乐性节目的安排策略。



由于娱乐性节目的提升情况非常好，我们对这一类型的节目继续细化研究：分为电视剧、综艺节目、电影、体育、其他；提取样本后发现娱乐性节目占到了总体样本的 69%



由上图可以看到电视剧类、综艺类的提升度最佳，即使是非热门节目也有提升；而四大类节目外的其他部分则出现了负影响。其中体育类节目仅抽取到热门的部分。

### iii ) 提升度与主要播放段个数：

通过节目提升度相对于播放段个数的柱状图 ( 提升度=该段数下提升度总和/该段数频次 ), 可以看到主要的播放段集中在 1 个时比较有利于提升频道收视率, 不宜过多, 而这与前文中关于播放段的讨论也是基本符合一致的。

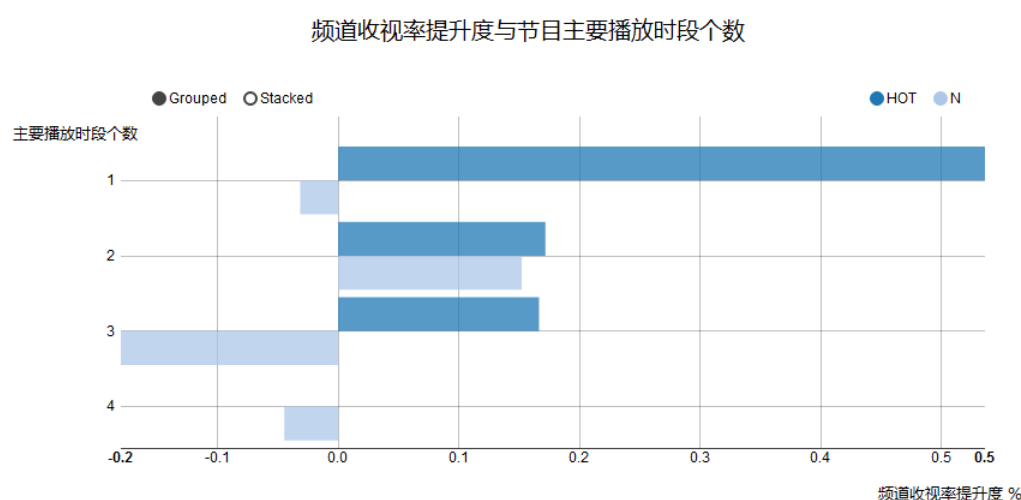


图 53

### iv ) 提升度与频道类型

通过节目提升度相对于频道类型的柱状图 ( 提升度=该频道类型下提升度总和/该频道类型频次 ), 考虑到是春节期间, 央视频道主要都在播放春节节目, 与去掉春节节目的收视率比有显著提高是很正常的。但可以看到地方台的提升又明显优于江苏本地台( 更多当地的针对性, 如新闻类节目 ) 的提升度。说明对这类大范围下的热门节目更适于非针对性频道选择。

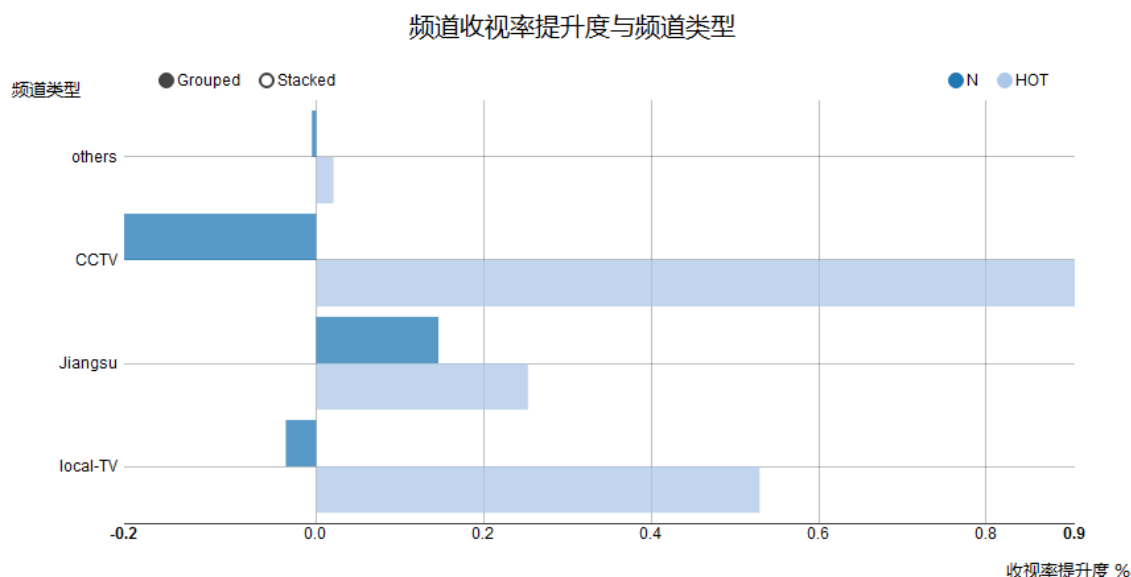


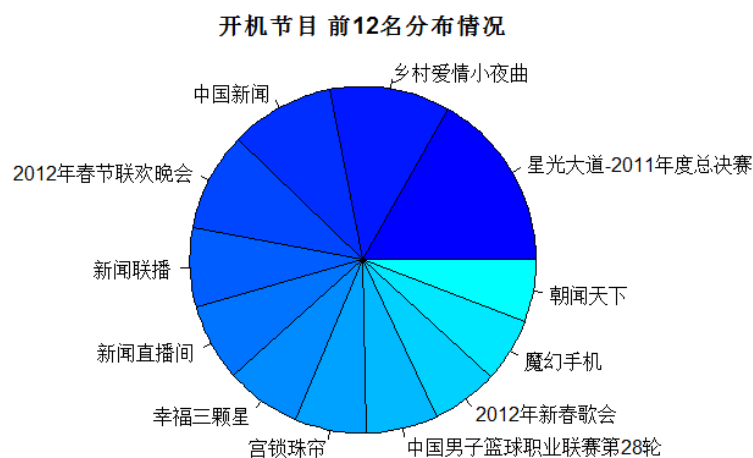
图 54

结论：

- i) 热门节目与频道优势节目基本保持一致，适合各类频道选择
- ii) 即使是热门节目，播放时间段并不一定要选在最黄金的时间段，若优势不是非常明显，下午连续剧时间或是零点左右，仍然有家庭主妇、退休老人及放假学生、夜猫族这些收视人群，略避开黄金段不一定有非常好的收视率，但能得到较好的提升；
- iii) 娱乐性质的节目最能提升频道收视率，而服务类最弱；而娱乐类节目中，又以综艺类节目与电视剧类型的节目提升最佳。
- iv) 主要播放段集中在 1 个时比较有利于提升收视率；
- v) 选择热门节目提升收视率的方法更适用于江苏地区以外的地方卫视（针对性较小），而央视因为有春节节目的干扰，对日常性其他节目仍有待分开讨论。

## 4. 开机情况

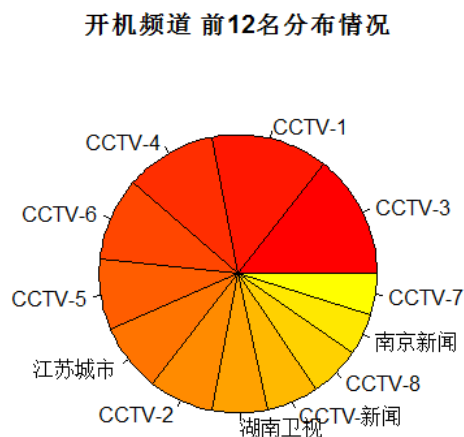
### 4.1. 节目



图表 45

开机节目排名和收视率排名几乎一致,可以看到收视率和开机出现次数呈正相关的关系。

### 4.2. 频道



图表 46

对比各个频道的收视率我们不难发现,开机排名前12的频道和收视率排名几乎一致。

但是江苏城市和南京新闻例外,收视率排名在50开外,不考虑每天第一次开机的情况,可以发现这两个频道的相对开机频率下降,



而且开机锁定频道排名中也位列前 10，说明这两个地方节目虽然观众群体不够大，但是存在一定的收视稳定性，大多数观众都是准点开机进入这两个频道观看特定的节目。

## 5. 用户收视节目预测

利用收视节目的分层属性（之前将节目类型已经做过分类和标记）我们将用户的节目收视分为两类：节目类别（定义为分层-1）和节目（定义为分层-2）。不同的节目类别会包含很多不同的节目，但是节目类别的数量明显小于节目数据量。因此，节目类别之间的转化是相对稳定的，所以在分层-1 可以利用马尔科夫模型对用户可能收看的节目类型进行过滤；预测用户在  $t$  时刻可能浏览的节目类别之后，可以在分层-2 利用贝叶斯定理并结合用户在  $t-1$  的状态预测用户在  $t$  时刻可能收看的节目。

定义

$Record_i = \{(category_{i1}, program_{i1}), (category_{i2}, program_{i2}), \dots, (category_{in}, program_{in})\}$ : 表示用户收看节目类别和具体节目的情况， $category_{il}$  表示第  $l$  个网页的类别。

$S$ ：节目相似度矩阵；  $P$ ：马尔科夫转换矩阵；  $R$ ：相关性矩阵。

模型构建

### 1、节目相似度矩阵的建立

当收看第  $l$  个节目时，可以从 128557 个用户中获得  $l \times 4$  矩阵（节目

个数不满 4 个补 0 ), 该矩阵中列向量表示节目类别 i 被第 m 个用户收看, 则记 1, 否则记 0.

用欧几里得相似度计算两个不同节目类别的相似度并进行标准化:

$$ND(V_i^{\text{col}}, V_j^{\text{col}}) = 1 - \sqrt{\frac{\sum_{k=1}^m (V_{ki}^{\text{col}}, V_{kj}^{\text{col}})^2}{m}} \quad (1)$$

则相似度矩阵 S 通过 (1) 计算得到:

$$S = \begin{bmatrix} S_{11} & \cdots & S_{1k} \\ \square & & \square \\ S_{k1} & \cdots & S_{kk} \end{bmatrix} \quad (2)$$

## 2、马尔科夫转换矩阵的建立

$$P = \begin{pmatrix} P_{11} & \cdots & P_{1k} \\ \square & & \square \\ P_{k1} & \cdots & P_{kk} \end{pmatrix}, P_{ij} = \frac{\text{Number}(i, j)}{\sum_{j=1}^k \text{TotalNumber}(i, j)} \quad (3)$$

$P_{ij}$  代表从一个节目类型到另一个节目类型的概率。(3) 中分母表示节目类型从 i 转换到全部 4 个类型的总次数, 分子表示所有类别转换到类别 j 的总次数。

## 3、相关性矩阵的建立

$$R^n = \begin{bmatrix} R_{11}^n & \cdots & R_{1k}^n \\ \square & & \square \\ R_{k1}^n & \cdots & R_{kk}^n \end{bmatrix}, R_{ij}^n = S_{ij} \cdot P_{ij}^n \quad (4)$$

矩阵中每个元素都表示两个类别直接的相关程度。

4、利用贝叶斯定理计算下一个最可能收看的节目 program , 并获得最可能被收看的前几个节目, 预测的节目集合将作为分层-2 的输出结果。

$$P(\text{Programbi}|\text{programa})=P(\text{programa}|\text{programbi})P(\text{programbi})/P(\text{Programa})$$

数据运行

取每个用户 l=3 的收视数据记录 ,根据第一个节目和第二个节目推算出第三个可能收视节目的预测集合, 计算准确率。

S1=

1.0000000 0.1546545 0.4391081 0.4467742

0.1546545 1.0000000 0.1687588 0.1739122

0.4391081 0.1687588 1.0000000 0.4685758

0.4467742 0.1739122 0.4685758 1.0000000

S2=

1.0000000 0.1718408 0.4508255 0.4604259

0.1718408 1.0000000 0.1580619 0.1642923

0.4508255 0.1580619 1.0000000 0.4395071

0.4604259 0.1642923 0.4395071 1.0000000

P=

0.2089254 0.4756377 0.1394451 0.1759918

0.07767827 0.68664265 0.11685919 0.11881989

0.1512596 0.1064879 0.2844400 0.4578126

0.1350257 0.4927908 0.1371408 0.2350427

$R1 = S1 * P^2 =$

0.06277511 0.4126226 0.06548619 0.14987231

0.01981642 0.5506128 0.03358780 0.06388664

0.05160777 0.3040359 0.10056190 0.25146196

0.04950368 0.4312263 0.06778070 0.16974829

$R2 = S2 * P =$

0.3526347 0.8685320 0.3509023 0.5110232

0.1596722 0.8661700 0.2083118 0.2600409

0.3170712 0.6460345 0.4260507 0.6592380

0.3104619 0.8713990 0.3455574 0.5368069

从 R 矩阵得到：-1 层节目类型集合为：{ 2、3 }

因此，在该数据集中，收看节目数为 3 的用户，在前两个收看行为已知的情况下文艺娱乐和教育类被收看的可能性最大。

## 6、用户细分研究（基于用户行为和兴趣的聚类）

用户收视数据属 整理

属性名称	数据类型
1、 用户收看平均时长	整数
2、 Session 个数	整数

3、 Session 平均时长	整数
4、 节目个数	整数
5、 频道个数	整数
6、 一天的收视开始时间	1-24
7、 频道种类	1-4
8、 节目种类	1-4

## 7、模糊评价模型

目的：欲根据频道收视率提升度部分的讨论，对电视台的收益情况进行综合评估。由于数据来源是江苏地区，遂对中央电视台、江苏电视台进行讨论分析。

模型建立：

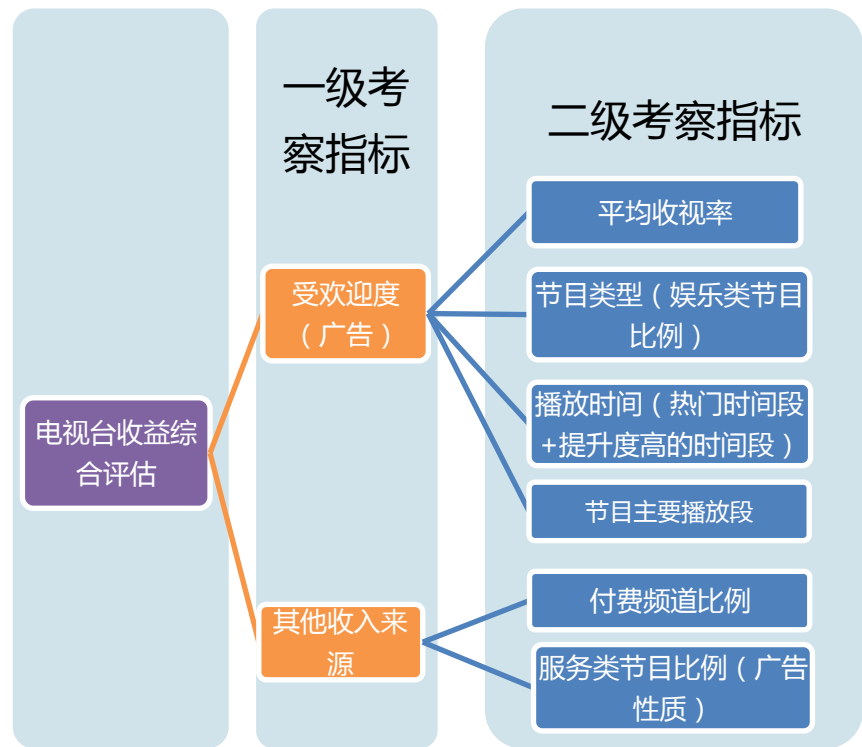
### 1) 数据选取：

电视台	类型	频道
中央电视台	普通	CCTV-1 (综合) CCTV-2 (经济) CCTV-3 (综艺)
		CCTV-4 (中文国际) CCTV-5 (体育) CCTV-6 (电影)
		CCTV-7 (军农) CCTV-8 (电视剧) CCTV-9 CCTV-10 (科学·教育)
		CCTV-11 (戏曲) CCTV-12 (社会与法) CCTV-新闻 CCTV-少儿
		CCTV-音乐 CCTV-高清 CCTV-中视购物
	付费	CCTV-风云剧场 CCTV-风云音乐 CCTV-风云足球 CCTV-第一剧场
		CCTV-世界地理 CCTV-怀旧剧场 CCTV-央视精品 CCTV-发现之旅
		CCTV-高尔夫·网球频道 CCTV-高清影视

江	普通	江苏卫视	江苏城市	江苏综艺	江苏影视	江苏少儿
苏		江苏体育	江苏公共	江苏教育		
电	付费	江苏靓妆				
台	海外	江苏国际				

注：中央电视台有 714780 条；江苏电视台有 262618 条

## 2) 评价指标：



## 3) 权值确定：

### i) 利用层次分析法确定二级指标权值：

根据 Saaty 等人提出的一致矩阵法构建判断矩阵，其中尺度表如下

( $u^1$  是受欢迎度的二级指标； $u^2$  是收入来源的二级指标)：

尺度 $a_{ij}$	含义
1	$u_m^i$ 与 $u_n^i$ 的影响相同

3	$u_m^i$ 比 $u_n^i$ 的影响稍强
5	$u_m^i$ 比 $u_n^i$ 的影响强
7	$u_m^i$ 比 $u_n^i$ 的影响明显的强
9	$u_m^i$ 比 $u_n^i$ 的影响绝对的强
2,4,6,8	$u_m^i$ 与 $u_n^i$ 的影响之比在上述两个相邻等级之间
1,1/2,1/3,...,1/9	$u_m^i$ 与 $u_n^i$ 的影响之比为上面 $a_{mn}$ 的互反数

分别得到受欢迎度与收入来源两项的二元对比较矩阵：

受欢迎度	平均收视率	节目类型	播放时间	主要播放段	权重 A
平均收视率	1	9/5	7/4	4	0.428
节目类型	5/9	1	4/5	5/3	0.211
播放时间	4/7	5/4	1	3/2	0.233
主要播放段	1/4	3/5	2/3	1	0.129

由于最大特征根  $\lambda_{\max}$  为 4.0255，一致性指标 CI=0.0085，CR<0.1，一致性检验通过，所以受欢迎度的权重即上所示。

其他收入来源仅两个因素，显然可以得到付费频道比例与服务类节目比例为 0.727，0.273

ii) 建立评价尺度（三人打分）

	央视得分	解释	江苏得分	解释
平均收视率	5、4、5	8.41%	2、3、2	5.43%
	总体平均为 6.71%			
节目类型	3、3、3	娱乐类节目比例 57.14%	5、5、4	81.30%
	娱乐类节目收视率提升度最大，总体平均为 56.27%			

<b>播放时间</b>	4、3、3	优势时间段占 55.67%	4、3、3	55.35%
	11:00-21 : 00 间收视率及其提升度较高，对总体而言，比值是 53.94%			
<b>主要播放段</b>	4、3、4	加权平均为 1.55 个	5、4、5	加权平均为 1.37
	主要播放段为 1 时收视率提升度最大			
<b>付费频道比</b>	4、3、3	10/27=0.37	3、2、2	1/10=0.1
<b>服务节目比</b>	3、2、2	10.80%	2、1、1	6.94%
	服务类节目比越高，则收益越好，其总体平均为 16.41%			

注：5-优秀、4-良好、3-一般、2-较差、1-糟糕

### iii ) 对二级指标进行模糊综合评价

由上表，得到归一化的隶属矩阵：

央视：

	<b>5</b>	<b>4</b>	<b>3</b>	<b>2</b>	<b>1</b>
<b>平均收视率</b>	2/3	1/3	0	0	0
<b>节目类型</b>	1	0	0	0	0
<b>播放时间</b>	0	1/3	2/3	0	0
<b>主要播放段</b>	0	2/3	1/3	0	0
<b>付费频道比</b>	0	1/3	2/3	0	0
<b>服务节目比</b>	0	0	1/3	2/3	0

江苏电台：

	<b>5</b>	<b>4</b>	<b>3</b>	<b>2</b>	<b>1</b>
--	----------	----------	----------	----------	----------



平均收视率	0	0	1/3	2/3	0
节目类型	2/3	1/3	0	0	0
播放时间	0	1/3	2/3	0	0
主要播放段	2/3	1/3	0	0	0
付费频道比	0	0	1/3	2/3	0
服务节目比	0	0	0	1/3	2/3

对满意度进行二级模糊综合评价：

央视  $b_1 = (0.4963, 0.3063, 0.1983, 0, 0)$

$b_2 = (0, 0.2423, 0.5756, 0.182, 0)$

江苏电台  $b_1 = (0.2266, 0.191, 0.298, 0.2853, 0)$

$b_2 = (0, 0, 0.2423, 0.5756, 0.182)$

iv) 对一级指标进行模糊综合评价

最后分别得到中央电视台，江苏电视台的关于受欢迎度及其他收入来源的模糊评价打分：

	受欢迎度（广告商的倾向）	其他收入来源	综合
中央电视台	4.30	3.06	3.8
江苏电视台	3.36	2.06	2.84

评价：通过模糊综合评价模型，不论是广告商的青睐程度，还是电视台的其他收入来源，中央电视台的得分上都比江苏电视台高了将近一个得分点；但相对其他收入来源一项，二者在广告商的青睐程度上差距更小一些。