

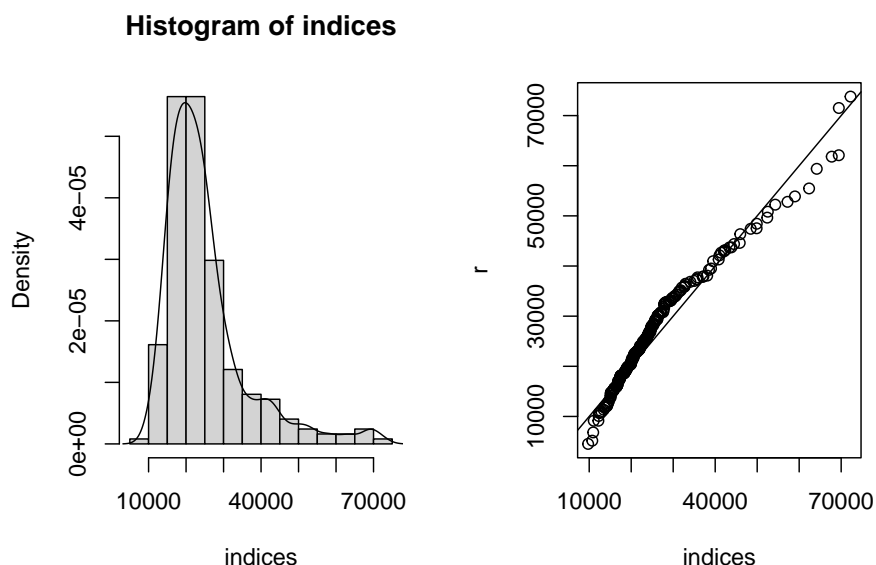
双十一数据分析

尚尔淦 彭小雨

目录

1	关于数据处理的说明	2
2	双十一参与购物人群数据分析	3
2.1	预期消费与实际消费	3
2.2	所购商品的种类	11
2.3	为谁而购物	16
2.4	购物平台	21
2.5	省份	31
2.6	18-25 岁全日制学生 (以下简称目标群体)	32
2.7	总体性别与目标群体性别 (以下简称第二类讨论群体)	37
3	双十一未参与购物人群数据分析	44
3.1	由收入情况分析	44
3.2	由评价分析	46
3.3	由男女比例 ~ 消费评价分析	47
3.4	目标人群分析	49
3.5	职业	51
4	整体总结	51
5	双十一购物推荐	53
6	作业分工	53

1 关于数据处理的说明



在数据清洗的过程中，我们要对“收入”进行具体化，为了更合理的给出每个收入数据的具体数值，我们采取如下方法：，我们首先从中国国家统计局网站中调出近多年各省市的年平均收入数据，绘制核密度估计图。我们可以从数据的核密度估计图中看到分布呈现单峰，右侧明显有一拐点，因此我们猜测，数据分布大致服从伽马分布，事实上，对于伽马分布函数的核 $f(x) = x^{\alpha-1}e^{-\beta x}$ ，我们分别求一阶和二阶导数

$$f'(x) = e^{-\beta x} x^{\alpha-2} (\alpha - 1 - \beta x)$$

$$f''(x) = x^{\alpha-3} e^{-\beta x} (\beta^2 x^2 - 2\beta(\alpha - 1)x + (\alpha - 1)(\alpha - 2))$$

得到极值点和拐点坐标为

$$x_1 = \frac{\alpha - 1}{\beta}$$

$$x_2 = \frac{\alpha - 1 + \sqrt{\alpha - 1}}{\beta}$$

我们若能从核密度估计图中读出极值点和拐点分别为 a, b 那么通过求解方

程我们可以得到 α, β 的表达式为

$$\alpha = 1 + \frac{a^2}{(b-a)^2} \quad \beta = \frac{a}{(b-a)^2}$$

, 同时我们利用如下事实:

$$X \sim \Gamma(\alpha, \beta) \rightarrow \frac{X}{12} \sim \Gamma(\alpha, 12\beta)$$

但是尽管我们有如上理论结果, 但是实际操作中我们会发现在 $X \sim \Gamma(\alpha, 12\beta)$ 时, 生成 20000 元以上的随机数的概率近似为零 (系统不可能产生这样的随机数), 因此我们考虑到尾部概率为无穷小, 在 $X > 20000$ 的条件下 (条件概率) 我们姑且可以认为他服从均匀分布, 我们为了使数据尽可能紧凑, 我们设置均匀分布为 $U(20000, 40000)$

注: 此外, 根据我们对于数据分析的需求, 我们将数据进行分词, 赋值, 引入新数据等处理, 具体处理方式见代码 “数据处理” 部分

结论 1(收入分布). 被分析数据中 “收入” $income \sim \Gamma(\alpha, \beta)$

2 双十一参与购物人群数据分析

在初步完成对数据的清洗后, 我们形成了一个初步的数据分析思路, 即我们以下将对于表格的每列进行初步分析

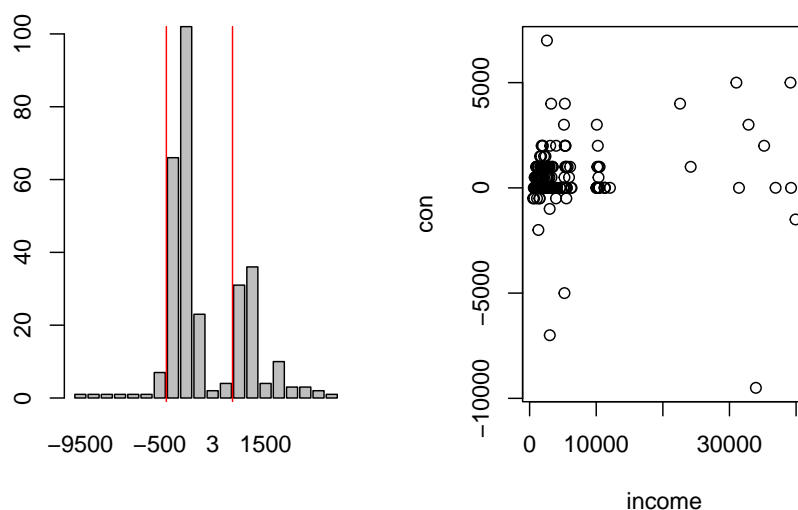
2.1 预期消费与实际消费

2.1.1 初步分析

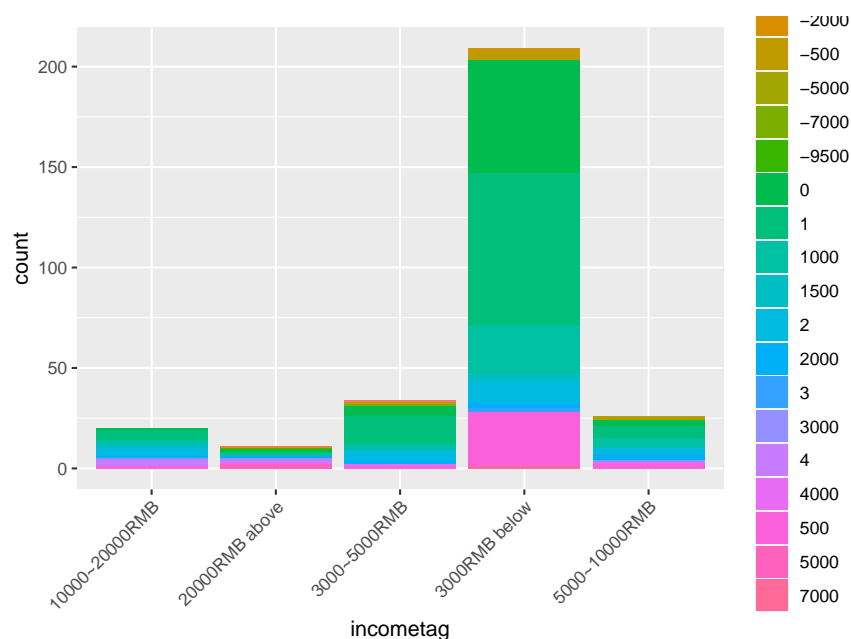
结论 2.1. 我们由柱状图中可以明显发现大致有半数的人群能够做到预期消费与实际消费在同一区间内, 但是剩余半数人群的实际消费几乎均超过了预期消费

2.1.2 预期消费 & 实际消费与收入的关系

由于在日常生活中，我们可以体会到收入从某种程度上决定着人们的消费水平，于是该部分我们结合收入进行分析



结论 2.2. 承接上述分析，我们在柱状图中加入红线将人群分为三类，由进一步改进的图中我们进一步发现：绝大多数预算和实际差不多的人群收入较低，而收入较高的人群中的多数都在双十一活动中参与了消费。随即，我们对收入与实际消费是否超出预期消费做出了散点图，由散点图我们可以发现：收入高低和实际消费是否超出预算的程度并不具有明显的相关性，我们可以由此推断出收入高低并不对人们的理财观念产生决定性影响

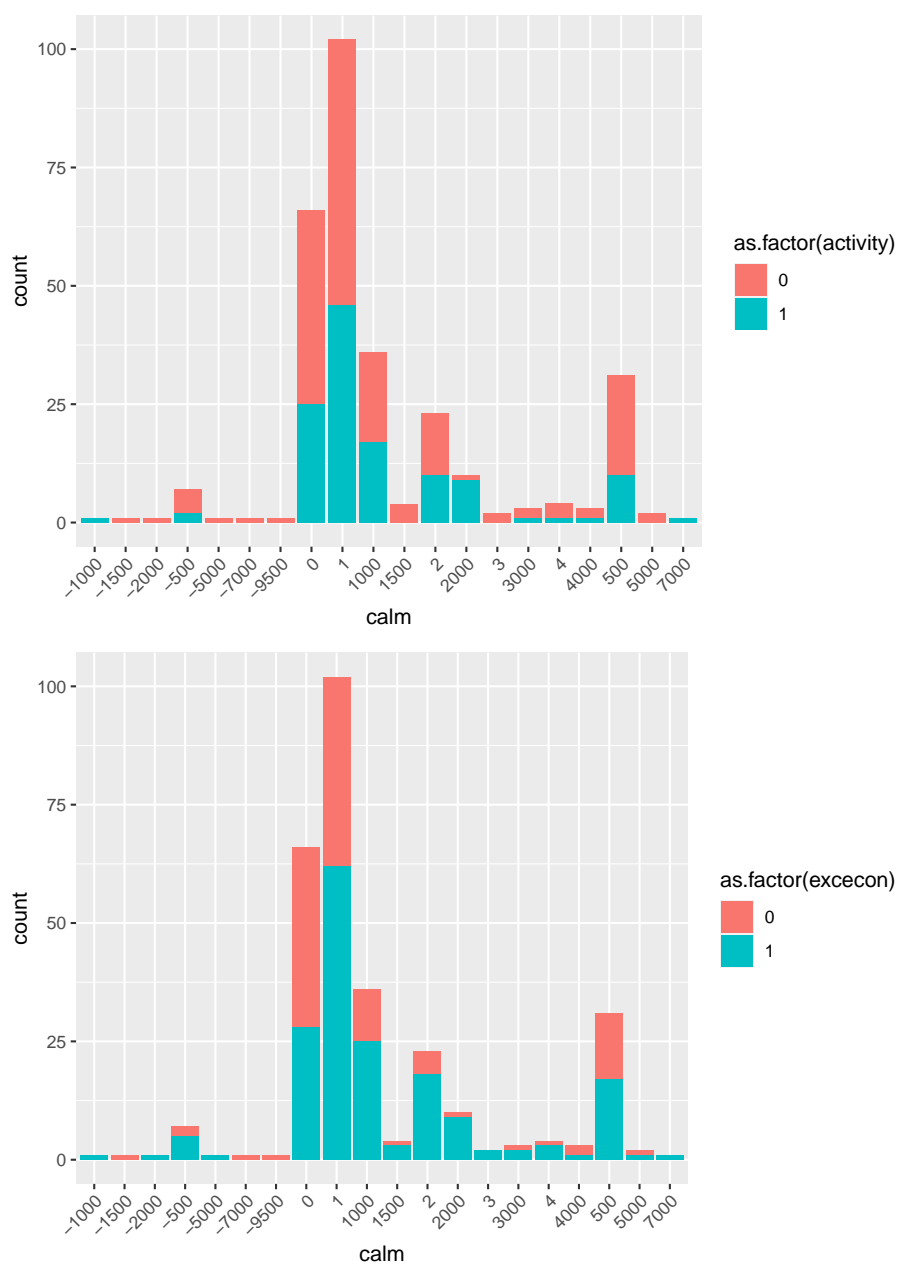


结论 2.3.

：我们进一步画出收入与消费是否理性的柱状图，从该图中我们可以清晰地看出高收入人群大多预期消费与实际消费重合，但是低收入人群中超前消费的人群占了较大比例，但是在全民剁手的潮流中仍然有较多数人能够做到理性消费，这或许可以说明剁手消费的潮流只是商家营造出的一种为了刺激消费的虚拟的社会氛围

2.1.3 预期消费 & 实际消费与是否参与活动以及使用花呗 & 白条的关系

我们在生活中不难体会到，在我们的购买过程中，折扣活动与花呗 & 白条所造成的对我们的购物欲的影响是不容小觑的，因此我们以下通过数据分析判断我们的猜测是否正确



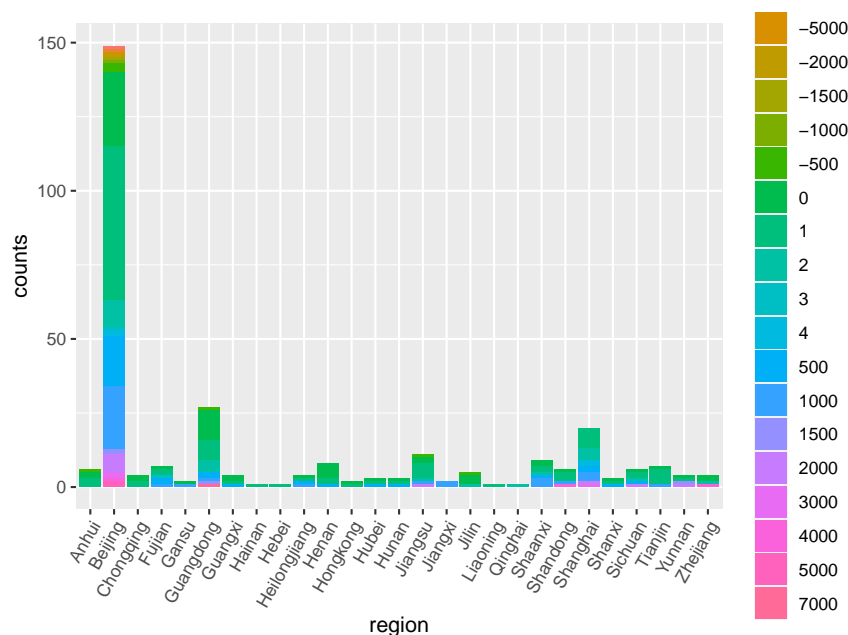
结论 2.4. 由第一个柱状图，我们可以看出对于能够控制住自己消费水平的人群（预期和实际在同一区间）对于是否参加活动并不是十分在意，参与与否呈现五五分的状态，但是我们发现实际消费小于预期消费的人

群多数都参加了活动，从某种程度上说这种折扣活动的目的是为了降低实际消费，同时实际消费较高的人群相比预期消费较高的人群，参加活动的比例增加很多，尤其是实际消费稍高于预期消费的人群（见图中横坐标 2000 处），这样的人群大多热衷于通过参加活动来降低实际消费，以防止实际消费过多超出预期，因此总体来讲尽管实际消费超出预期消费的人群大多参与活动但是仍然会导致实际超出预期（剁手），这说明商家的折扣活动对于刺激消费产生着一定的作用，而且这也表现出大众在消费过程中对于折扣活动存在较弱的抵制力

结论 2.5. 这个图片明显发现蓝色的比例（使用花呗和白条）较大，即大多数群众具有超前消费倾向，尤其是实际消费大于预期消费的人群使用花呗白条的频率更大，只有预期与实际近似的人群能够做到对花呗白条的使用进行均衡，但是也仍有半数的人超前消费。花呗白条的使用是近几年该开始兴起的支付方式，并一时间受到了众多年轻人的追捧，而我们的数据分析也恰恰印证了这种新兴的支付方式极大程度上起到了促进消费的作用，尤其对于消费不理智的群体的作用更加显著

2.1.4 预期消费 & 实际消费与地域之间的关系

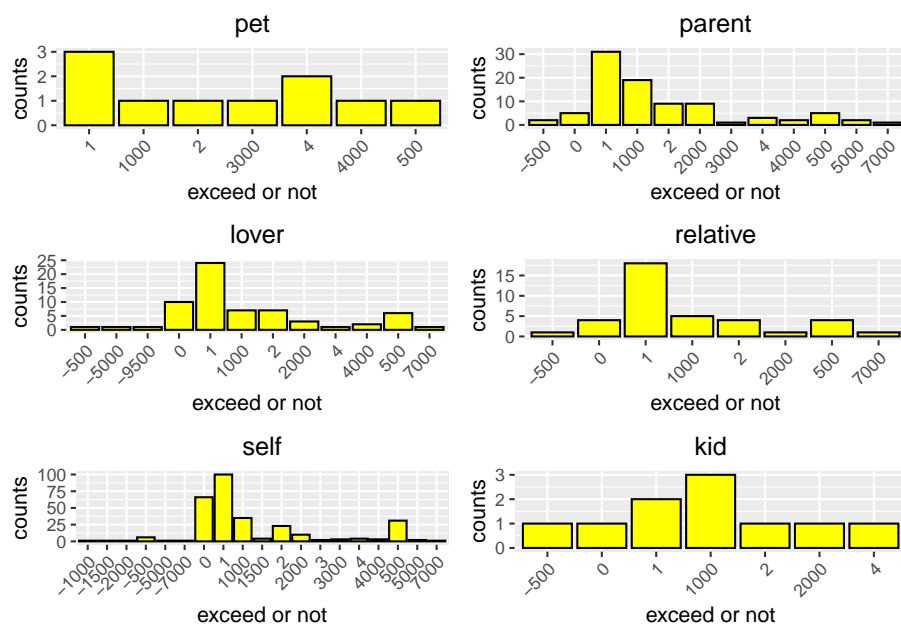
中国是一个拥有广袤地域的国家，而其辽阔的国土使得其各地区经济发展状况、乡土文化、收入水平和消费水平都有着显著不同，因此我们以下分析预期消费 & 实际消费与地域之间有怎样的关系



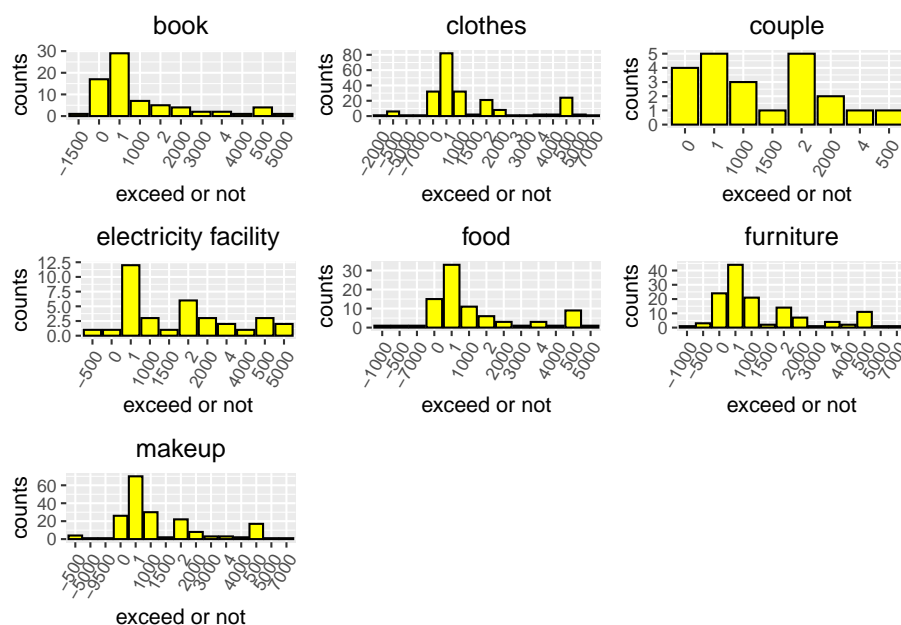
结论 2.6. 由柱状图我们能够发现：超前消费（实际消费大于预期消费）人群经常出现在如“北京”“上海”这样的发达城市以及如“浙江”“江苏”“山东”这样的沿海城市，其余城市人群的预期消费几乎均高于实际消费，尤其值得注意的是，北京三种消费人群（预期高于，等于，小于实际）层次分明，这说明超前消费的观念对各地域人群的影响程度有所不同，实际消费大于预期消费意味着在双十一活动结束后的一段时间内，超前消费的群体将会有较大的资金周转压力，而发达城市和沿海城市人群的高收入水平减轻了他们的心理负担从而更好的促进了超前消费观念的普及，这也让该地区群体成为了双十一消费的主力军

2.1.5 预期消费 & 实际消费与为谁消费 & 购买种类之间的关系

在购物中我们常常因为突然刷到的某种物品增加我们的购买力使得我们的实际花费显著增加，抑或是我们在购物中心血来潮想为亲人准备的惊喜使得我们的实际消费超出了我们的控制，我们以下就来分析为谁消费 & 购买种类是否会对我们的实际消费产生较大的影响



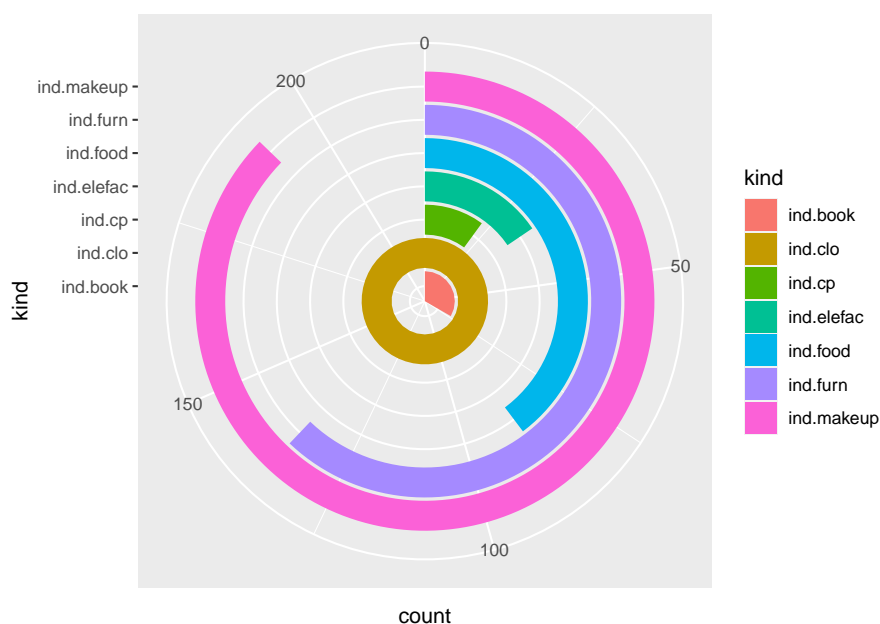
结论 2.7. 从图中我们不难看出小型类商品及日用品几乎不会成为我们实际消费大于预期消费的原因，而价格较高的电子类产品和家用电器成为了消费群体实际消费大于预期消费的主要因素，这说明所购产品的种类对预期消费 & 实际消费起到了一定的作用



结论 2.8. 我们可以由图像中发现，导致消费群体超前消费的两类主要群体为宠物和孩子，结合实际生活，我们可以推导出，消费群体一般在制定预期消费水平时往往会忽略宠物部分的花销，因而会造成宠物对超前消费有了一定的促进作用，而对于孩子而言，孩子本身的不定性性必然会导致实际与预期的不符，但是在排除这些不确定因素后，我们难发现绝大多数消费群体在消费中能较好地避免商品种类及人群对其消费的影响，因此为谁消费 & 购买种类并不会对预期消费 & 实际消费造成决定性的影响

2.2 所购商品的种类

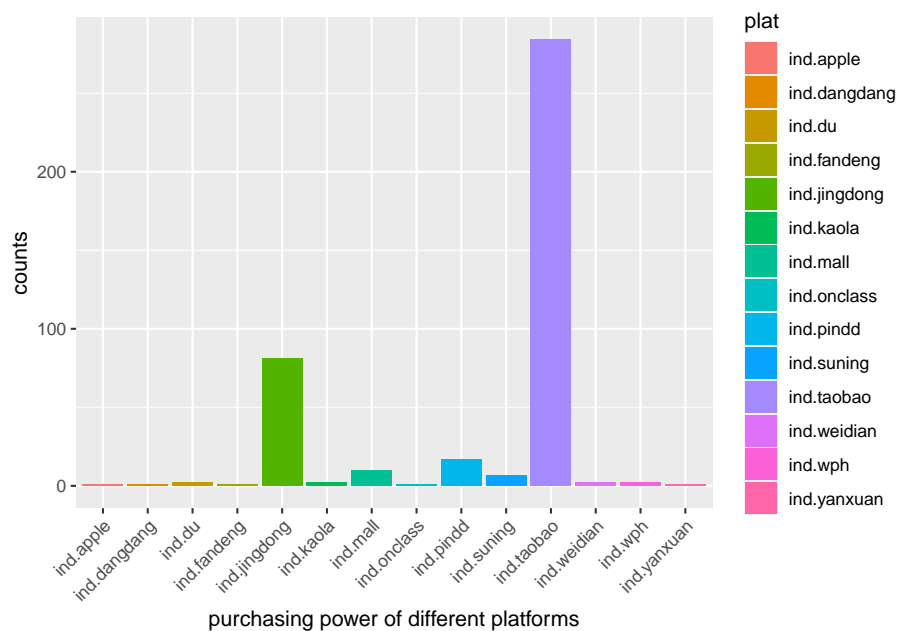
2.2.1 初步分析



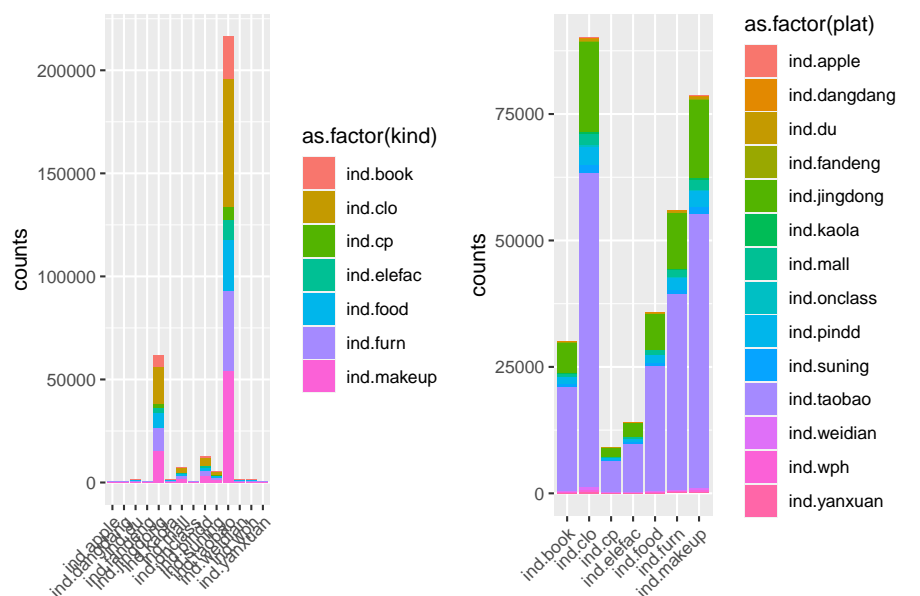
结论 3.1. 由图我们可以看出：服饰鞋靴的购买力度最大，美妆个护次之，这或许可以说明：当代社会现状下，人们对于物质方面的追求主要体现在穿搭方面，而且也从侧面说明了服饰和美妆需求量大，且更迭迅速(或许还可推断出购买人数中女性占据多数，此问题需待后续验证)，我们还可以从图像推断出，对于电子产品，家用电器等大型商品，大众普遍不会等到双十一再行购物，而对于消耗类，奢侈类产品，大众则更倾向于获取一定的价格优势

2.2.2 所购商品种类与平台的关系

电商平台与各大商场一样，或许都拥有着自己的优势销售商品，其不同商品之间的畅销程度一定也有着一定的不同，下面我们结合数据分析一下这两者之间是否确实存在着一定的关系



结论 3.2. 由柱状图可看出销量最好的平台为淘宝（天猫），京东次之，这或许会说明淘宝的促销力度大或是淘宝的平台可信度更高，以下我们做详细分析



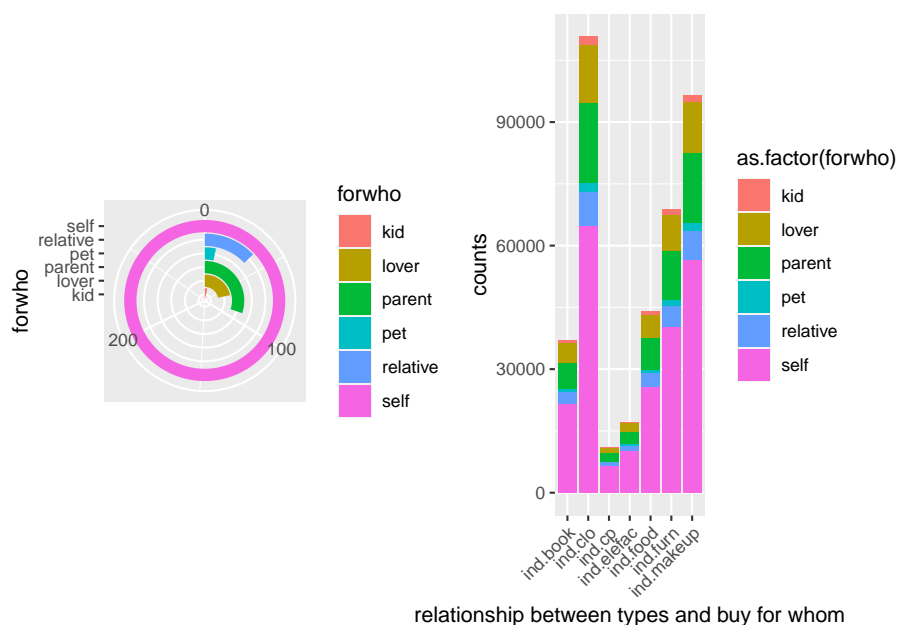
bution of different types in different platforms degree of different types of products on various platf

结论 3.3. 经由上一阶段讨论，我们该阶段只注重淘宝和京东两个平台种类购买量的研究：

- 我们可由柱状图发现，在淘宝平台上，服饰鞋靴的销量最好，美妆个护次之，家具百货次之，不同产品销量间存在明显的差异现象；但在京东平台上，服饰鞋靴和美妆个护的销量基本持平，家具百货和食品生鲜的销量同样基本持平，由此，我们可推断出，淘宝平台相较于京东的受众群体更为丰富，购买力度更大，但存在着产品种类的质量可信度，促销力度等的不均衡性，而相较于购买力度较小的京东，京东则更好的做到了不同产品之间各方面属性的均衡，降低了受众在购买不同种类的物品时受到的影响
- 我们还可以发现，在线上购买中，人们普遍更倾向于服饰和化妆品类的购买，而商场对于各类产品的销量则更为均衡，新兴的线上购物模式对服饰和化妆品的销售模式影响较大，但对于各类电子产品，家用电器及图书类物品，受众仍更倾向于传统的销售模式

2.2.3 所有商品种类与为谁而买的关系

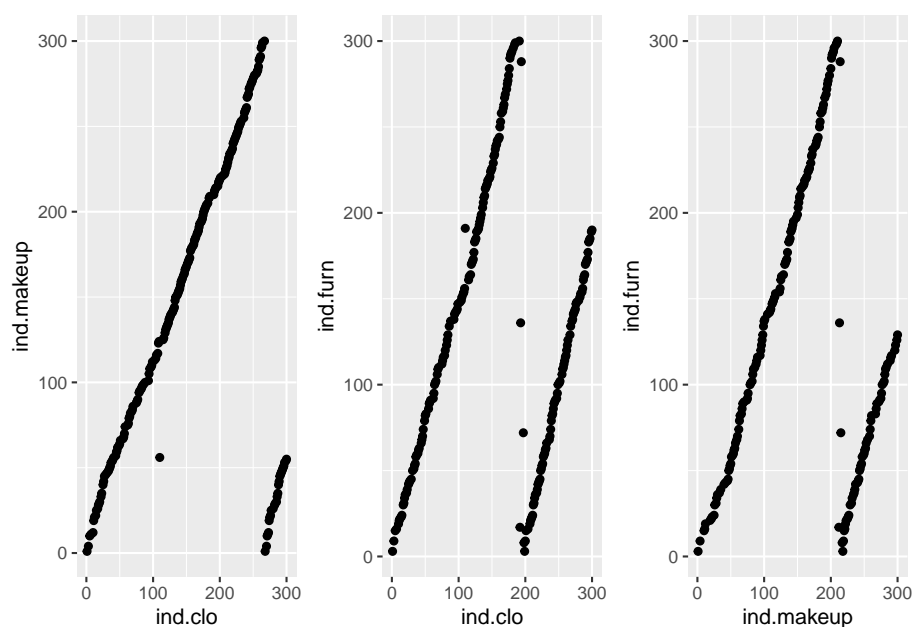
我们熟知不同的消费群体有着不同的侧重商品，于是我们考虑这种倾向性选择是否会带入到双十一的购物人群中呢，于是我们做了以下分析



结论 3.4. 由该图可看出，无论任一种类的商品，大众均更倾向于为自己购买，但我们还可以发现大众在为其他群体购物时会更倾向于购买衣物，化妆品等小型物品，更进一步，对于电子产品等大众更倾向于为亲人购物，此处需注意一个问题：我们的样本大多在 18-25 岁，他们大概率还没有孩子，因而我们根据目前样本并无法准确分析出在全国的双十一购物潮流中，大众是否具有为孩子购买商品的倾向

2.2.4 所购商品种类内部的关系

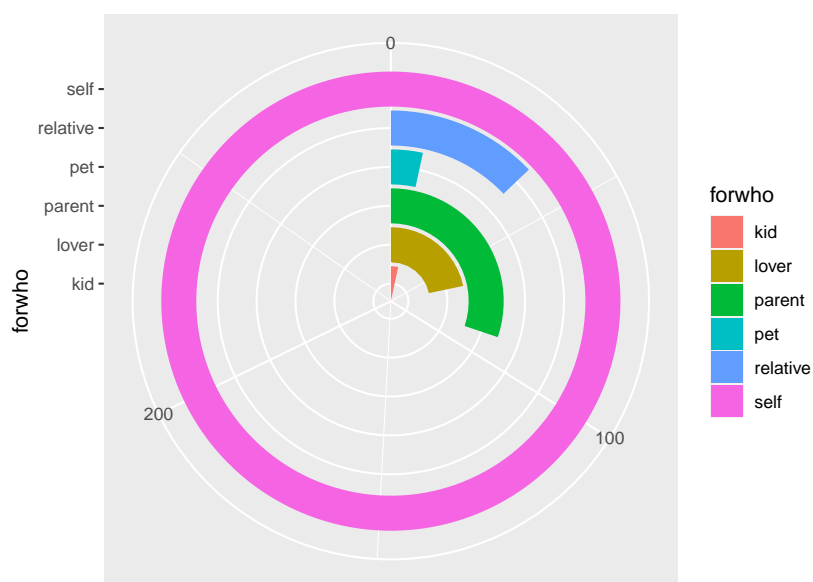
随着亚文化现象的兴起，我们不难发现每一种爱好都有了自己的社交圈，并且其所牵涉到的其他附属物品也随之得到了一定的增长，那双十一消费群体对于商品的选择是否也具有类似的特点呢，我们以下进行分析



结论 3.5. 由于购买化妆品, 服饰和家电的人数相对较多, 适合进一步进行种类内部分析, 否则对于购买量较少的与购买量较多的商品, 由于其数量的极度不对等性, 我们无法做进一步分析, 故我们分析它们三者之间的内部关系: 由散点图我们发现购买化妆品与购买服饰间存在较强的线性关系, 化妆品与家电次之, 服饰与家电更次, 由此, 我们可以推断出: 服饰与化妆品的需求量基本呈正相关性, 而其余种类内部之间联系则不存在该种关系

2.3 为谁而购物

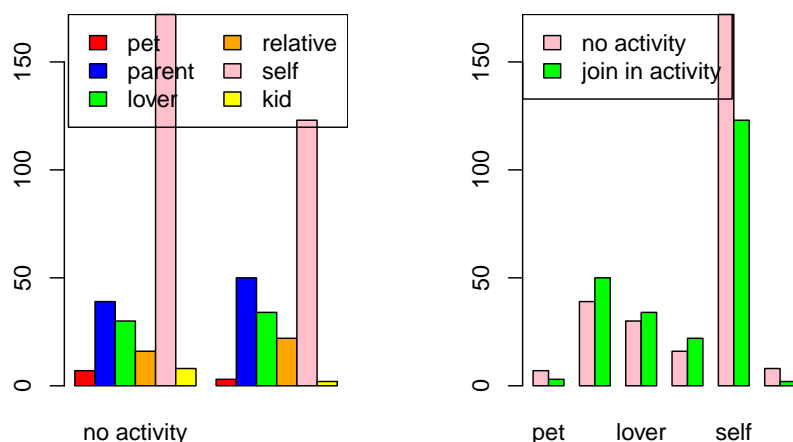
2.3.1 初步分析



结论 4.1. 由饼图我们可以明显发现，双十一购物人群中为自己购物的消费者占据了绝大多数，为父母消费的次之，为亲戚消费的再次

2.3.2 为谁而购物与折扣活动的关系

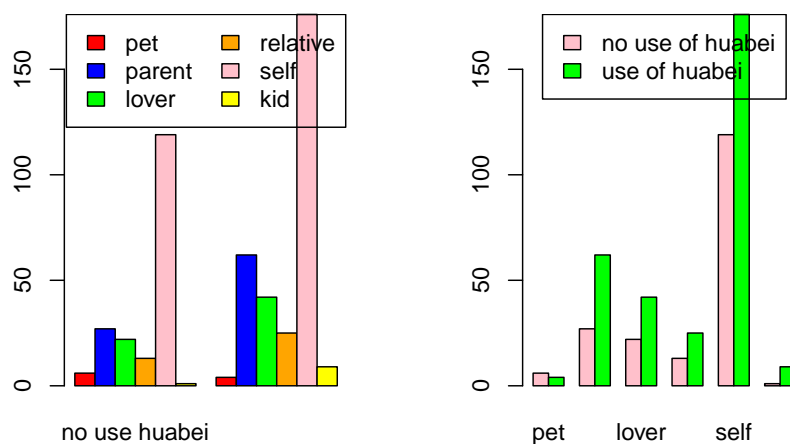
我们在购买商品时经常会发现，父母对于我们所期望购买的商品通常会尽量满足，而对于自己却秉持着能省则省的态度，下面我们希望通过数据分析观察出折扣是否影响着我们对于为谁购物的选择



结论 4.2. 上图可以发现无论是否参加活动，人们大多是为自己购物，为父母和伴侣购物的人数次之，人们在购物过程中很少考虑宠物以及孩子，参加活动与不参加的人数分配近似相同。我们换一种作图方式，得到了更为有价值的信息为为孩子购物的人群中不参加活动的人数较多，给人一种为孩子购物“无关紧要”的感觉，这或许恰恰应证了我们在数据分析时的猜想，即在为孩子选购商品时，考虑折扣活动的因素会被大大削减，消费群体在对折扣与满足孩子的需求方面会更倾向于选择后者

2.3.3 为谁而购物与是否使用花呗 & 白条的关系

据我们在日常生活中的观察不难发现，大众在消费时通常会优先选择信用卡、花呗和白条等透支的消费致富方式，而这种选择是否会影响到为谁购物就是我们接下来想要分析的内容

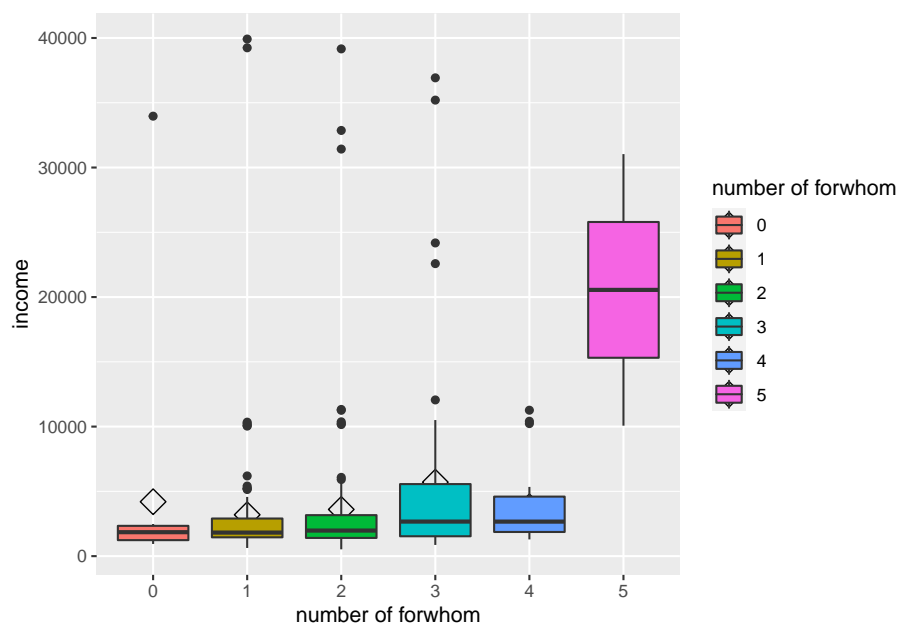


第二幅图的横坐标自左向右依次为:pet,parent,lover,relative,self,kid

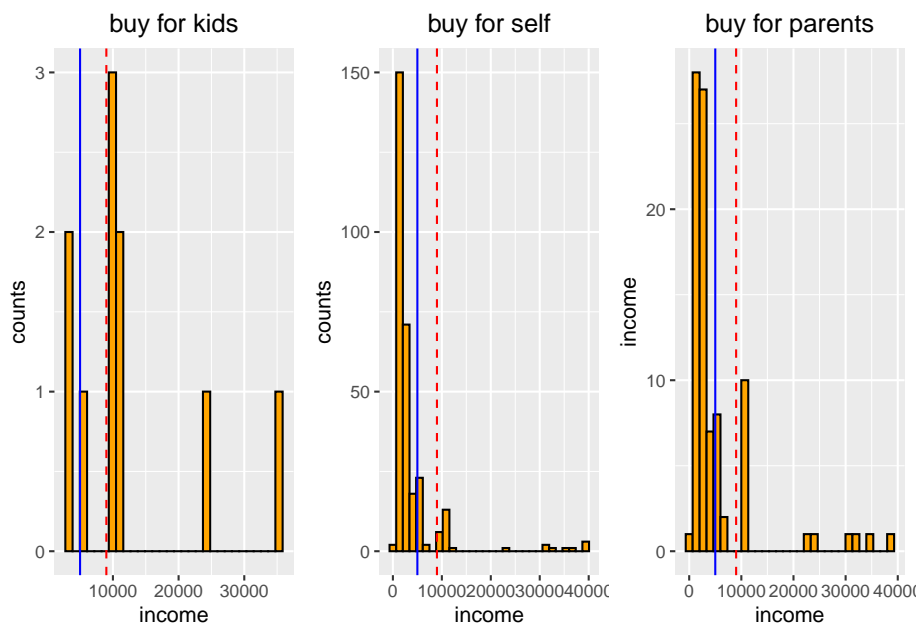
结论 4.3. 对于以上两幅图，我们可以发现无论为任一群体购买商品，使用花呗和白条的人数都要比不使用的更多，同时为自己购物而使用花呗和白条的人数占有绝对优势，次之是为了父母购物，如今的购物人群自我意识较强，而且透支式的支付方式并不对为谁购物产生选择性，人们潜意识在任何购物场景中都会优先选择该种支付方式

2.3.4 为谁而购物与收入的关系

我们已经了解到收入会对购物产生一定的影响，于是我们思考收入高的群体是否会为更多的群体购买商品



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



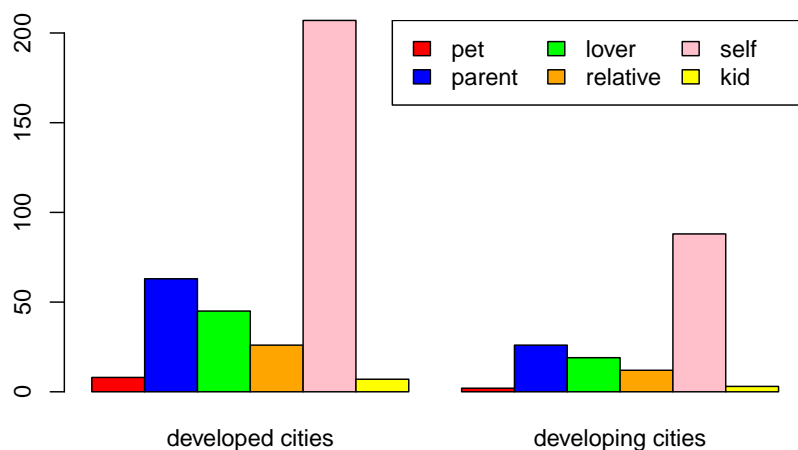
在之前的数据分析中我们察觉到为自己和父母购物的人群是占据了绝大多数，我们不免由图产生疑问：孩子究竟扮演了什么样的角色呢？我们下面的结论会给出回答

结论 4.4. 综上所述，从上两个图来看，能够发现低收入人群更倾向于为自己和父母消费，从而忽略到孩子的消费中，这符合我们“高收入人群更有能力支持孩子的开支”的直觉.

- 从图中我们可以明显看出来为五种人群进行购物的买家收入水平远高于其他人群，同时随着人群种类的增加，收入水平也在缓慢增加，我们再对“离群值”进行观察得出，不为任何人购买的离群值明显少于为 3,4,5 类人群购物的人群的离群值的个数，这也从侧面说明：收入高的人群不仅有能力也真正会为较庞大群体来进行购物，于是我们产生思考，那些收入高的人群到底喜欢为什么样的人（或物）进行购物.
- 对于孩子对消费的影响: 有半数为孩子消费的人收入水平超过了 10000 元，这似乎也预示着高收入人群有更多的经历去照顾孩子，为孩子消费，而同时我们还需注意到，我们的样本中绝大多数群体为 18-25 岁的全日制学生，这部分群体的收入较低，同时也无需承担照顾孩子的工作，因此我们无法得出低收入人群对孩子的消费是否一定会有所减弱以及他们对孩子的照顾是否也一定会有所减弱

2.3.5 为谁而购物与地区的关系

由于中国各地区经济发展水平及群众收入都有着一定的差别，那么我们以下分析地区是否会直接对消费者在双十一购物中为谁购物产生影响

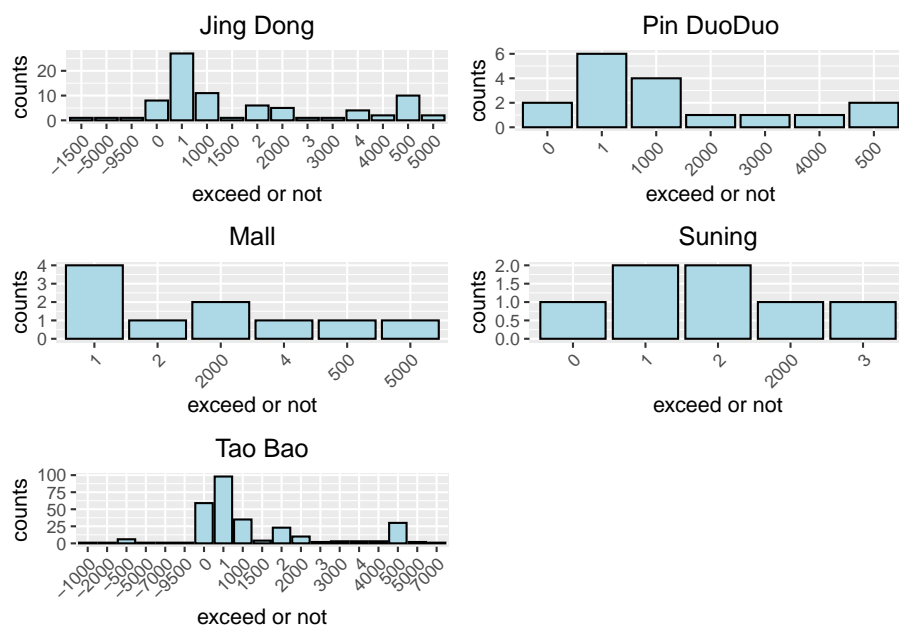


结论 4.5. 事实上，尽管发达城市购买力相较欠发达城市要强一些，但是购买需求（为何种人群购买）二者大致是相同的，不同城市人群的“生存需求”是类似的，对为谁而购物这个因素在地区维度上并不产生较大的差异

2.4 购物平台

2.4.1 初步分析

我们先对各平台的消费人数进行观察，发现大家消费的平台大多集中在京东，拼多多，淘宝等，因此我们绘图只考虑以上平台（关于其他平台的人数观察可以参照提交的代码中的注释段）

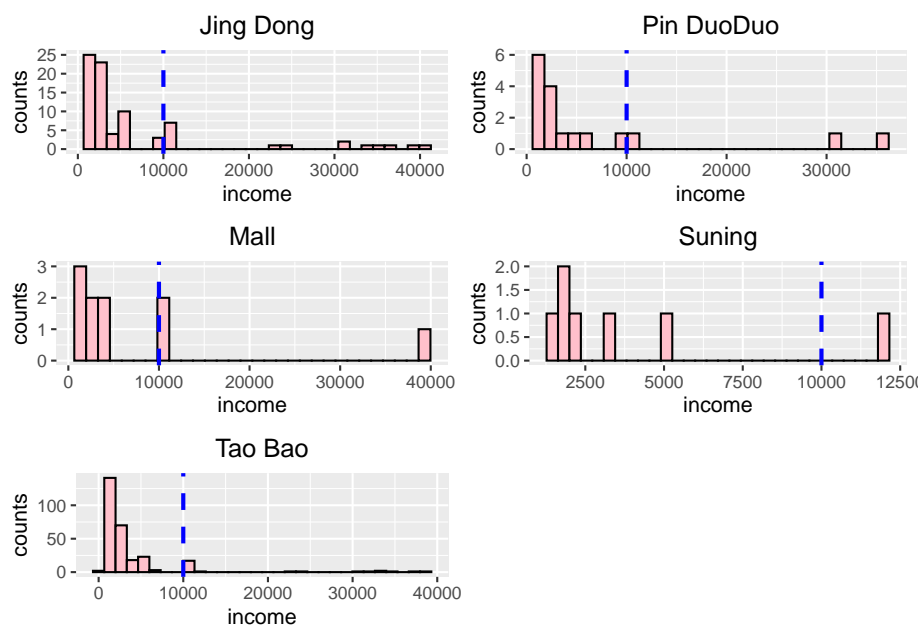


结论 5.1(一个有趣的发现). 尽管在拼多多消费的人群比较少, 但是这些人几乎全都参与了“剁手”, 拼多多真的是诱惑人家拼单搞得如火如荼, 而我们不得否认拼多多平台的拼单活动确实吸引了一部分消费者的注意力, 同时我们可以看出还有相当大的一部分人热衷于去淘宝(天猫)进行购物, 同时我们也看到了实体商场消费低靡, 同时对于京东和淘宝剁手的人的数量也比不剁手的人数多, 这说明京东和淘宝两家线上购物平台不仅有着庞大的受众, 还拥有着突出的促销手段

2.4.2 购物平台与收入的关系

线下实体商场常有各类档次划分, 那么线上购物平台是否也进行了类似划分呢, 我们以下根据数据进行分析

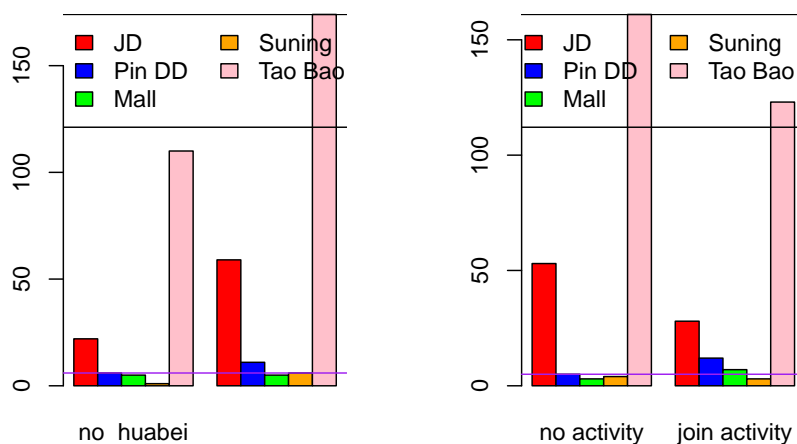
```
## `stat_bin()` using `bins = 30`. Pick better value with `binwidth`.
## `stat_bin()` using `bins = 30`. Pick better value with `binwidth`.
## `stat_bin()` using `bins = 30`. Pick better value with `binwidth`.
## `stat_bin()` using `bins = 30`. Pick better value with `binwidth`.
## `stat_bin()` using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



结论 5.2. 我们通过如下代码 (`\textbf{length(which(data_buy_taobaoincome > 10000)) > length(which(data_buy_jingdongincome > 10000))` 显示 `TRUE`) 可以看出高收入人群在淘宝购物的人数相对于在京东购物的人数多一小部分, 但是从低收入人群的角度, 更多人会选择去淘宝进行消费. 比较如上两个数值仍然可以发现在淘宝消费的人群普遍收入要低于在京东消费的人群, 这就印证了我们的上述猜想, 线上购物平台同线下购物平台一样, 对于消费层次有着大众默认的划分, 这另一方满也从侧面体现出京东平台的高质量商品或许会多余淘宝且其售后保障或许会比淘宝更加优越

2.4.3 购物平台与是否参加活动与是否使用花呗 & 白条的关系

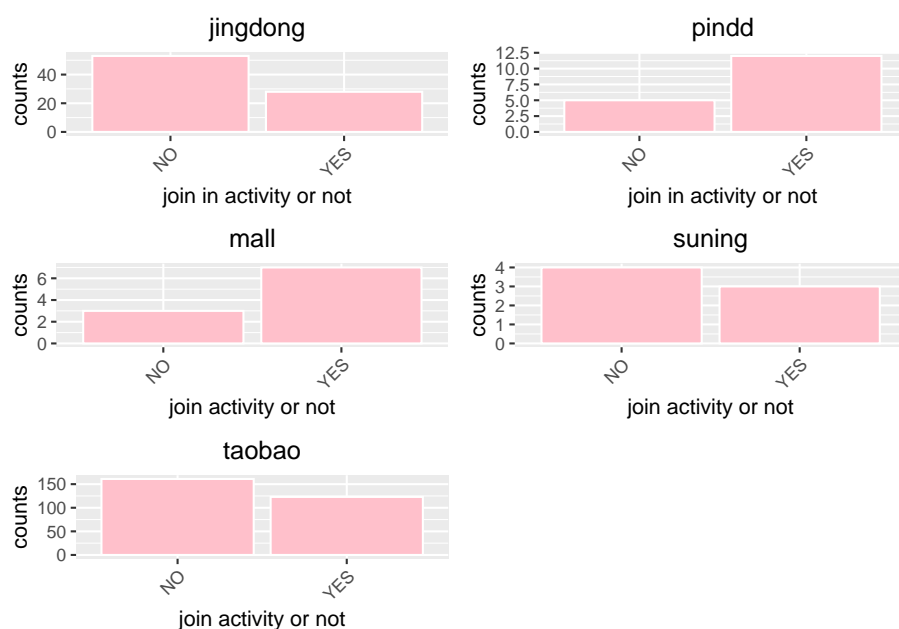
在传统的购物方式与购物理念中, 我们秉持着货比三家的购物理念, 那么在线上购物中这种理念是否依旧存在呢, 即我们以下分析平台的折扣活动与对花呗 & 白条的使用是否会对其自身的销售产生影响



结论 5.3. 见图中紫线能够发现，对于其他电子平台来讲使用花呗白条等工具的人的数量要高于不使用的人的数量，但是对于传统购物方式——商场来讲，是否使用这两种人群数量几乎是相等，这符合我们的直觉

2.4.4 平台购买量与是否参加活动的关系

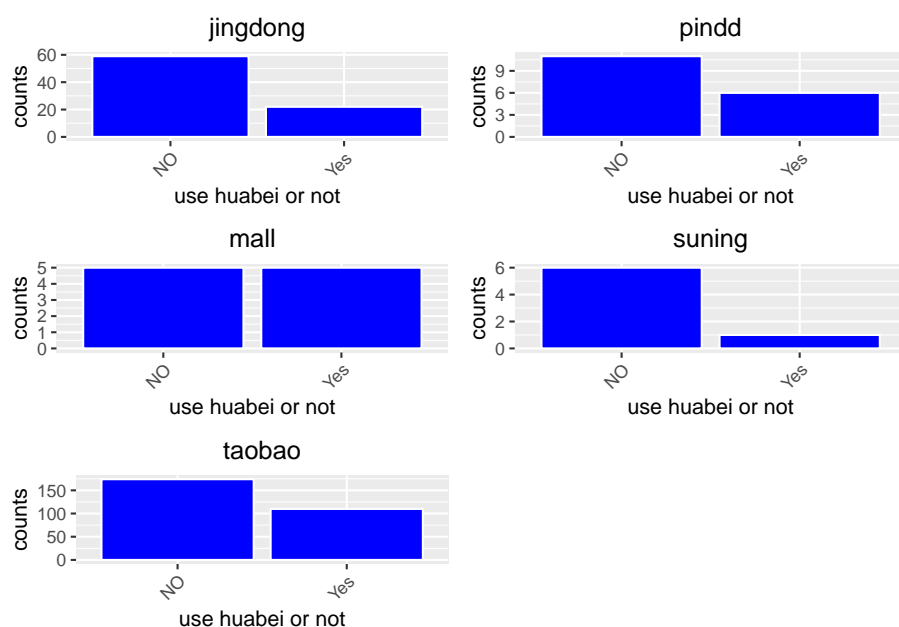
在线下购物中我们不难发现，在商场促销阶段购买的人数通常会远远超出非促销阶段的购物人数，以下我们分析这种局势是否还会持续到线上购物平台中



结论 5.4. 由平台购买量与是否参与活动的柱状图可以观察出：我们上述分析出购买量最大的两个平台：京东和淘宝的受众选择该平台购买商品受平台活动力度影响小，而其余购买量小的平台其受平台活动力度却相对较大，这说明：京东和淘宝两个线上购物平台已经成为了大众线上消费的首选，而且即使没有双十一的活动价，大众仍倾向于在这两个平台进行购物，这两个平台已初步建立起了公众线上购物的认可度

2.4.5 平台购买量与是否使用花呗 & 白条的关系

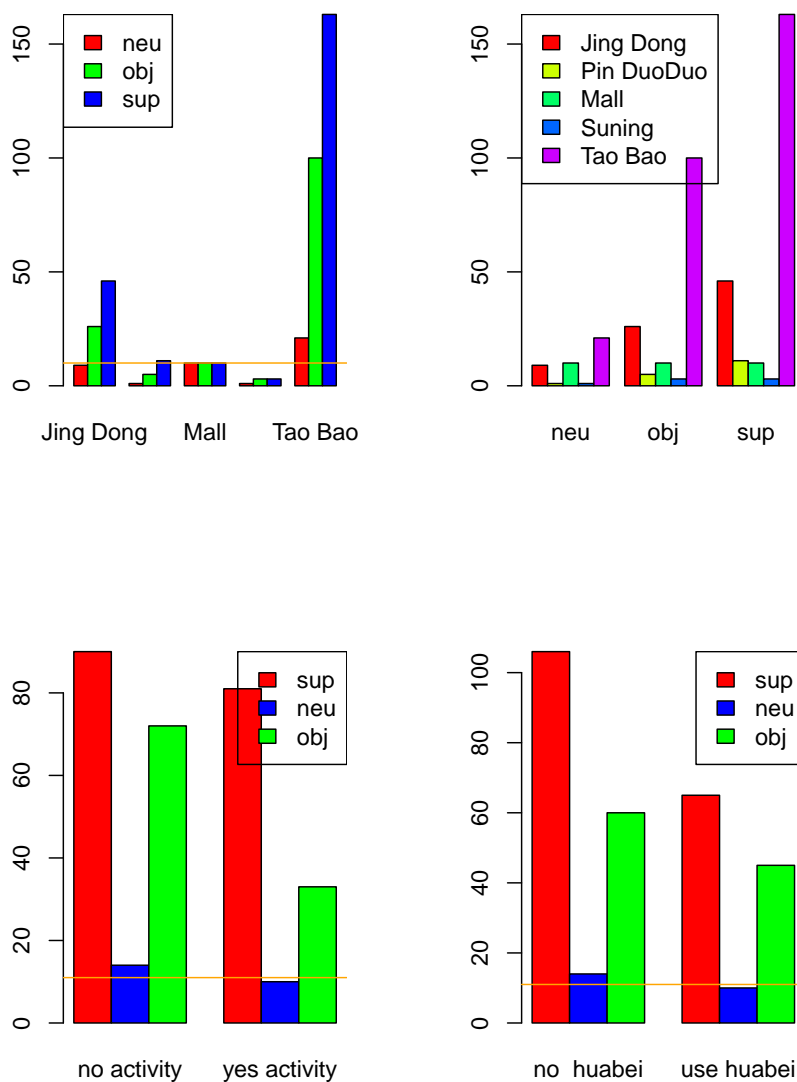
传统的购物模式中我们可以通过信用卡进行透支以为购物人群提供更便捷的支付方式，那么我们思考信用卡透支的进化版本花呗 & 白条是否会对平台购买量产生一定的影响



结论 5.5. 由平台购买量与是否参与活动的柱状图可以观察出：不同于是否参与活动对平台购买量的影响，是否可以使用花呗与平台购买量大小呈正相关趋势，说明花呗现阶段对大众如何购物，在哪购物已经起到了决定性因素，尤其是部分新兴的小型购物平台，如果不支持花呗作为支付方式，可能会最终失去本就稀少的购物人群，这也从侧面反映出传统购物方式逐渐失去大众喜爱的原因并凸显出当代社会环境下人们超前消费的理念进一步增强，而这也成为了各类线上平台促进消费的一种有力方式

2.4.6 购买平台与评价的关系

传统的购物模式中我们可以通过信用卡进行透支以为购物人群提供更便捷的支付方式，那么我们思考信用卡透支的进化版本花呗 & 白条是否会对平台购买量产生一定的影响



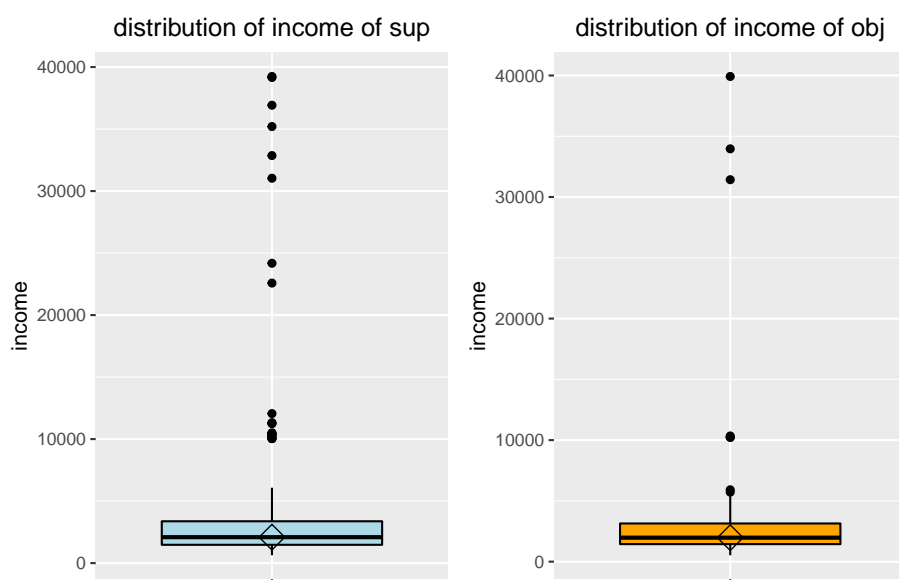
结论 5.6. 结合之前的几幅图表我们首先发现发现：

- 商场中购物的人三种态度完全相等，结合上之前图中商场购物的人参加活动与不参加活动；使用花呗白条与不使用，两种对立人群的数量

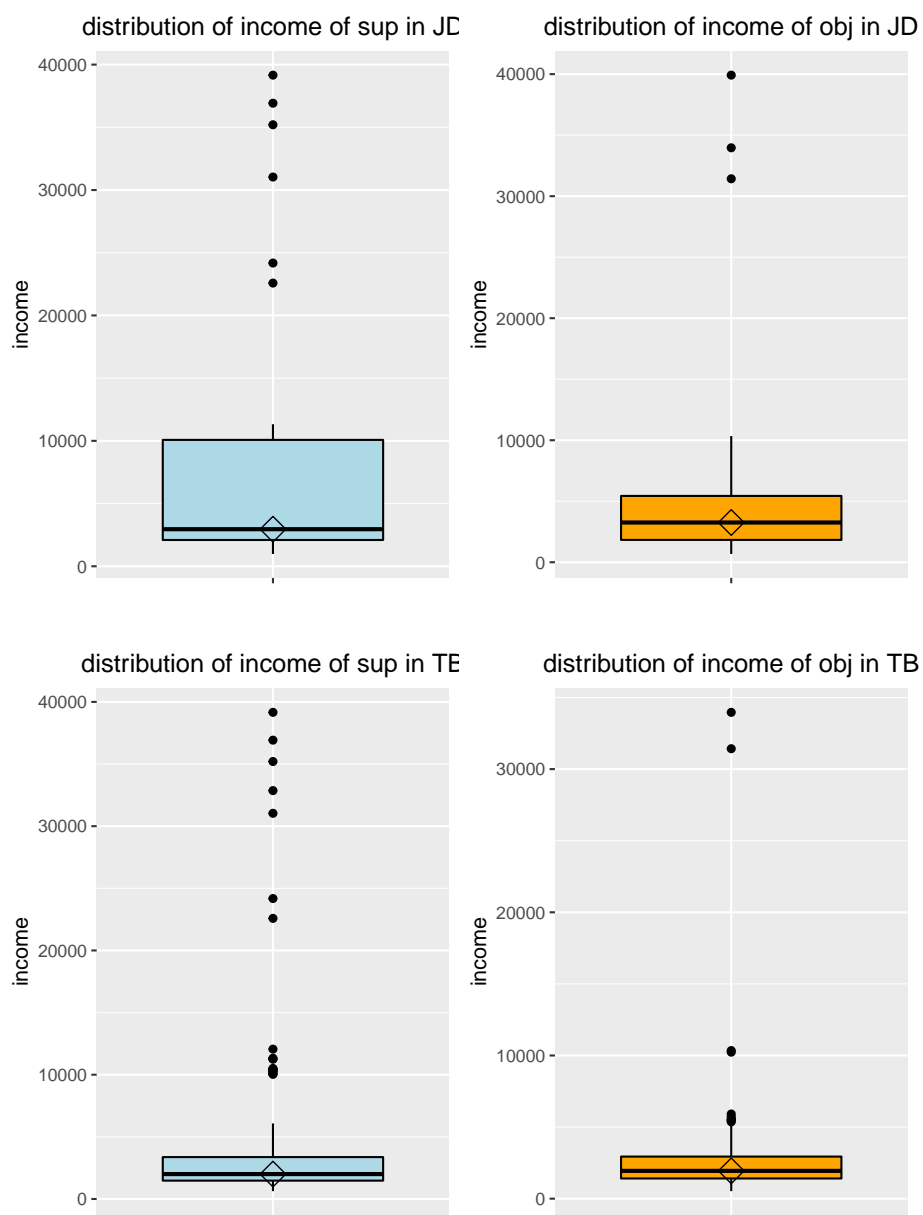
几乎相等，可以看出在互联网购物时代，商场的促销已经不能勾起人们的购物热情，反过来，可能人们也更愿意进行网上购物，对于实体商场不是十分关心

- 我们能够看出所有购物平台中持中立意见的人占少数，像京东和天猫这种互联网巨头平台中消费体验差的人数大概比支持的人数的一半多一点
- 我们还可以发现：淘宝的超前消费分布近似是以 0 为中心，使用花呗的人数比不使用的偏多，不参加活动的比参加活动的偏多（不参加活动为了消费当然要用花呗，这个很合理），淘宝平台中支持与反对的人比例近似为 8:5，这里我们进一步对评价与是否使用花呗 & 白条的关系可以发现，是否使用花呗 & 白条对于评价没有直接的影响，由于这里的数据表示购物人群消费后的评价，我们无法获知他们在消费前所接触到的评价的种类的相对情况，因此这或许说明评价具有一定的屏蔽作用，其与消费平台并无直接的互相决定的关系，受众不会因为平台的先导因素而在评价时产生偏颇

2.4.7 评价与收入的关系

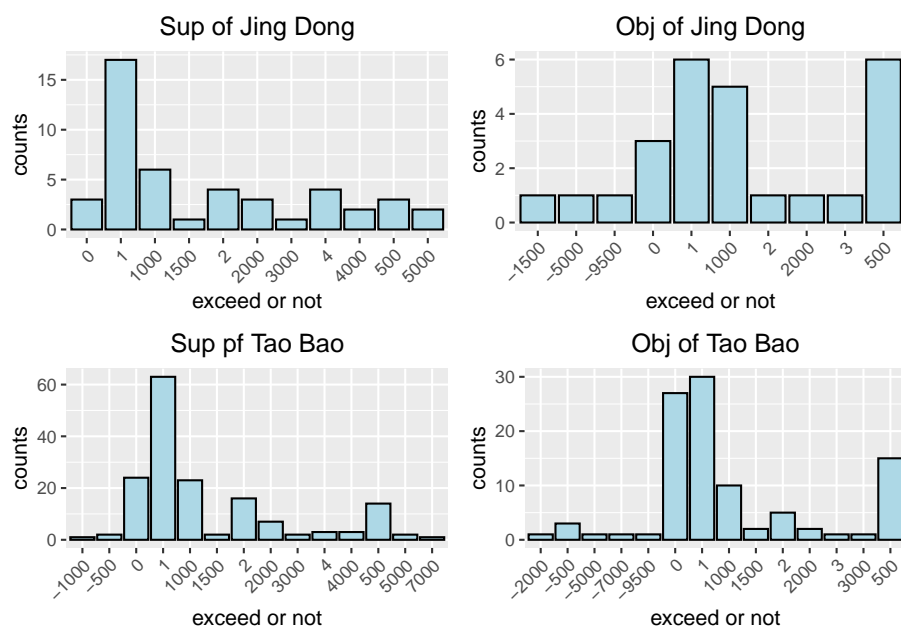


结论 5.7. 明显看出支持的人群高收入的比较多（有很多的离群值），同时也说明支持的人群收入水平分布区间比较广泛，相反也可以看出来反对的人群收入比较低，从而对于促销活动有反感情绪，这说明了人们的收入水平对于是否支持促销活动的推广有着直接的决定性因素



结论 5.8. 两大购物平台，总是支持的人的收入水平更高一点，我们之前也看到了京东的超前消费（实际与预期）呈现一点偏峰分布，这里京东支持的人的收入分布明显要比反对的人收入分布要广，我们由此推断可能是这个原因导致超前消费的水平不均匀，相比之下，淘宝两类人的收入区间大致相似（尽管有离群值但是箱线图相对于京东的更加端正）

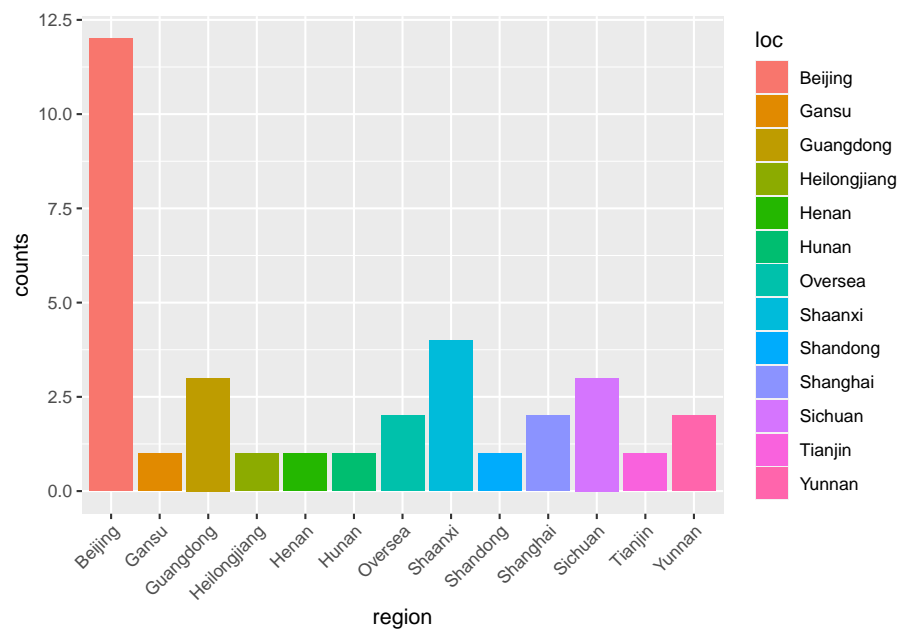
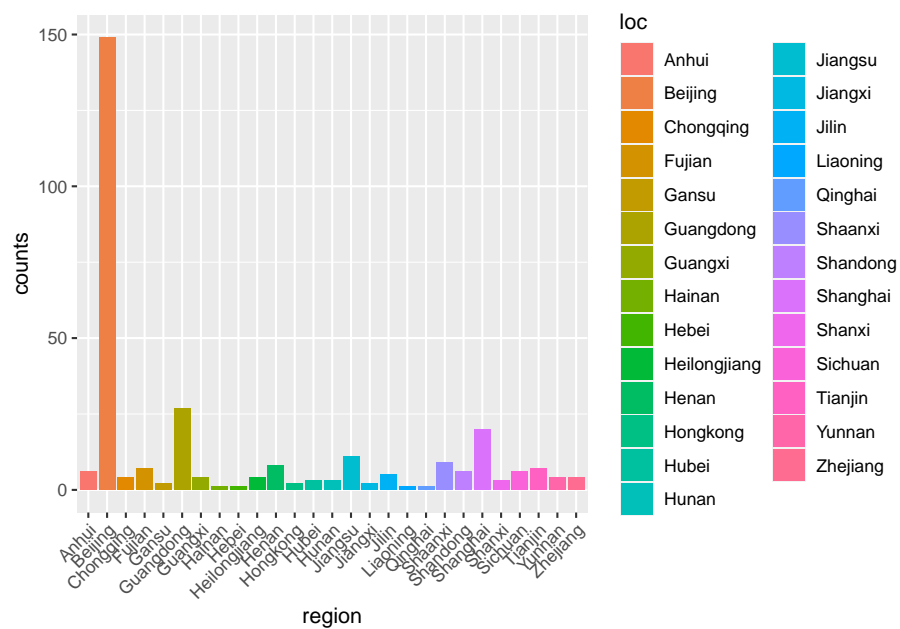
2.4.8 评价与是否超前消费的关系



结论 5.9. 由图中我们可以很明显地看出，选择京东平台进行消费的，尤其是支持双十一消费的人群剁手频率更高（这里几乎没有适度消费的都是预期低于实际），相反，淘宝中的人群要均衡得多，这也解释了为什么京东平台上超前消费的 bar 要偏峰而淘宝的峰在 0 附近，从上面图片中我们也可以看出在京东消费的人群的收入水平要比淘宝要高（要看相对比例），这些数据是自洽的。我们再从支持和反对人数的角度来分析，我们知道京东和淘宝分别的支持反对人数比例要近似，但是在京东消费的人数要比淘宝的小，因此在淘宝消费更能兼顾各项信息使得在淘宝消费的人群做到货比三家，理性消费，但是京东不一定如此，

因此不理性消费（剁手）现象还是更加普遍

2.5 省份

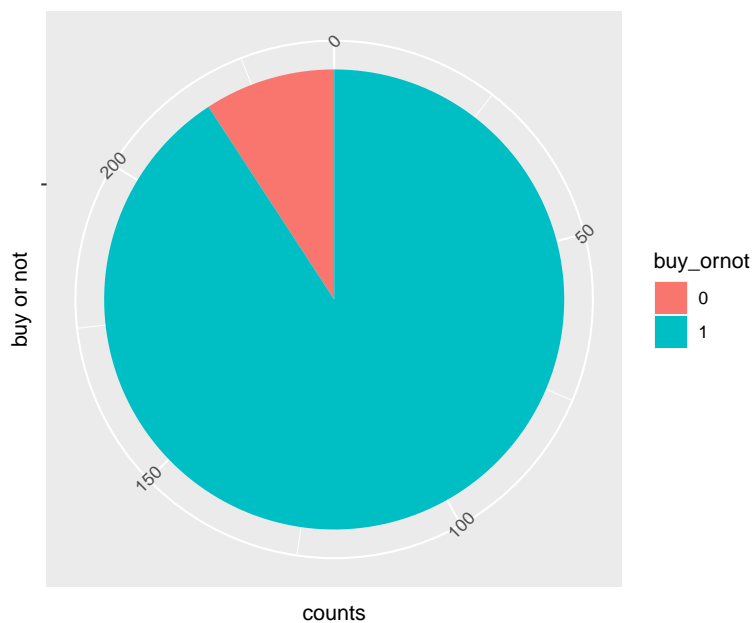




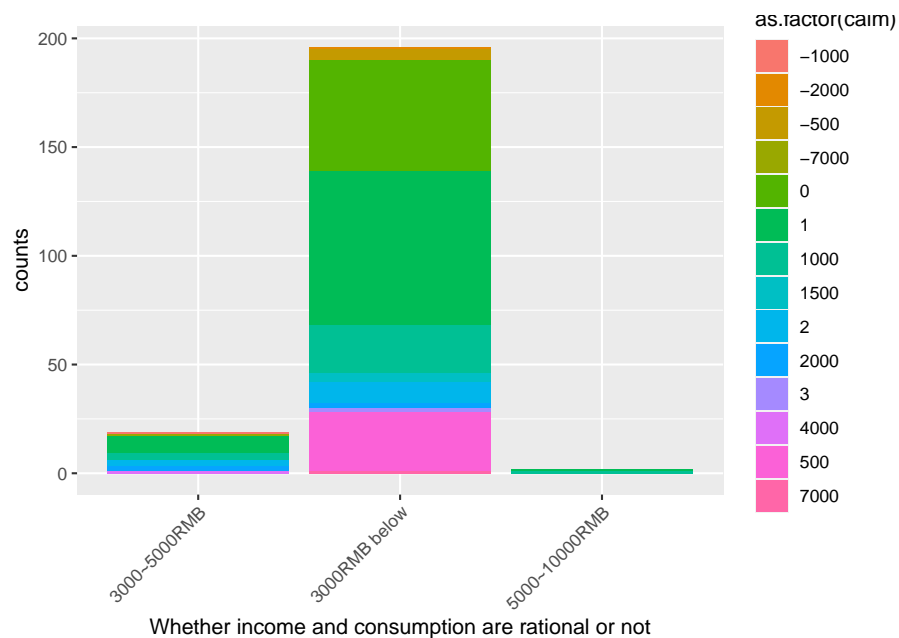
结论 6.1. 仅仅通过省份柱状图，我们无法判定购买力是否与省份直接相关，因为某些大省份比如：北京，广东等，在具有高参与购买人数的同时，未参与购买的人数同样很多，因此我们将省份与其他因素结合分析

2.6 18-25 岁全日制学生 (以下简称目标群体)

由下图我们不难发现参与双十一购物的 18-25 岁全日制学生为大多数，因此我们以下主要研究消费的 18-25 岁全日制学生的相关问题

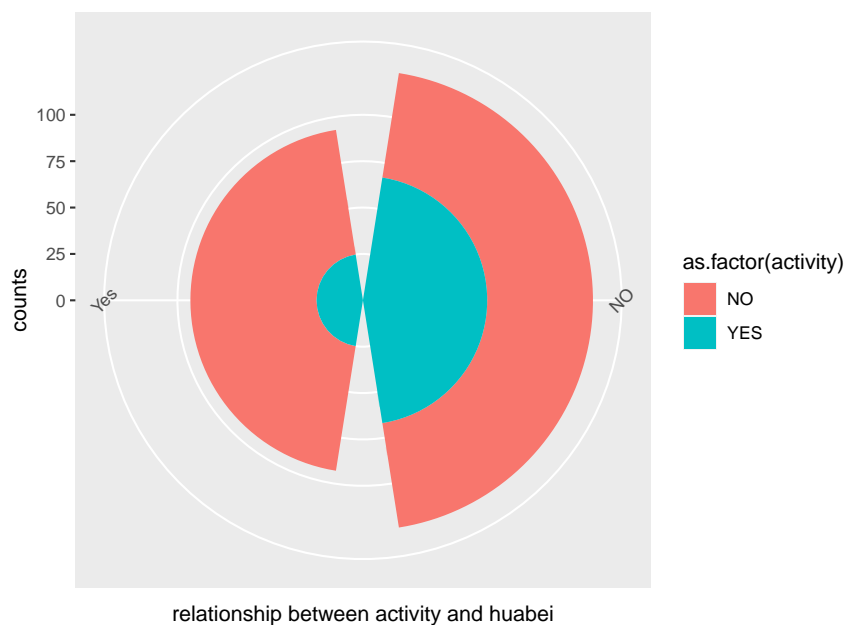


2.6.1 目标群体的收入于消费是否理智的关系



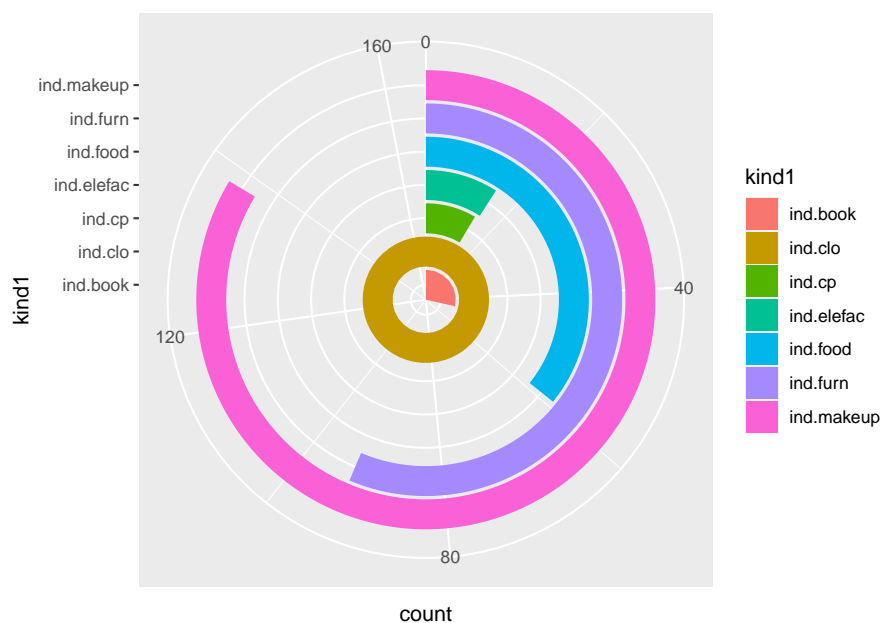
结论 7.1. 由此我们可以看出 18-25 岁全日制学生的收入主要在 3000 元以下且对于消费的理智程度较高，该群体中大部分可以做到理智消费但仍有数量客观的群体在消费过程中十分不理智，这表明 18-25 岁全日制学生中大部分受收入限制，可以实现较合理的理财观念，但是仍有部分样本在双十一购物中体现出极强的冲动心理

2.6.2 目标群体和是否参与活动以及是否使用花呗 & 白条的关系



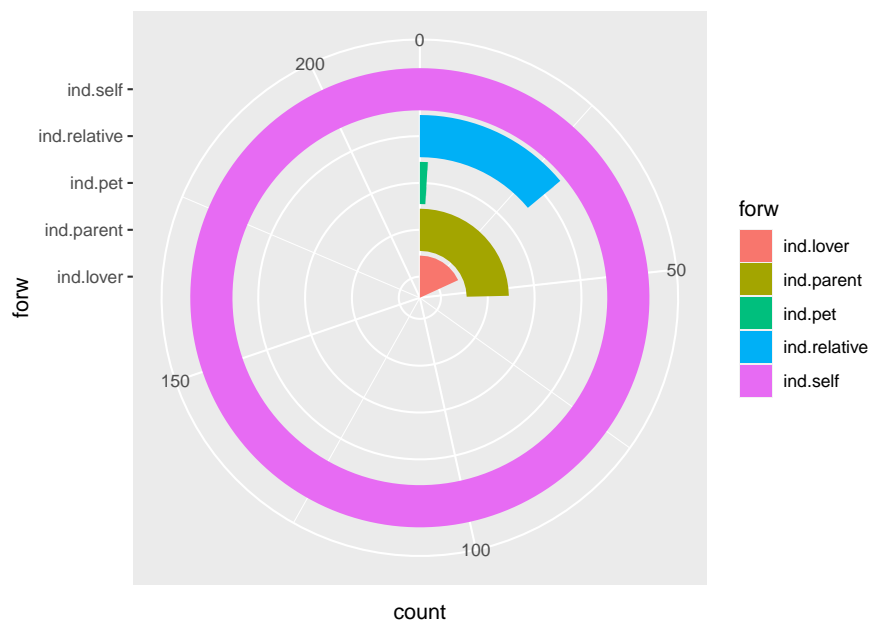
结论 7.2. 由以上玫瑰图我们可以看出 18-25 岁全日制学生群体中大部分在参与双十一购物中都使用了花呗进行超前消费，但参与活动的人数却相对较少，这一方面反映出当代大学生群体理财观念的缺失，同时也反映出大学生群体在购物时对于所购商品是否有活动并不关心，即活动价对于该群体诱惑力不强

2.6.3 目标群体与所售产品种类的关系



结论 7.4. 由图像我们不难看出该群体在购买商品种类上与全部样本没有太大区别，主要购买量仍集中在服饰与化妆品类购买物品上

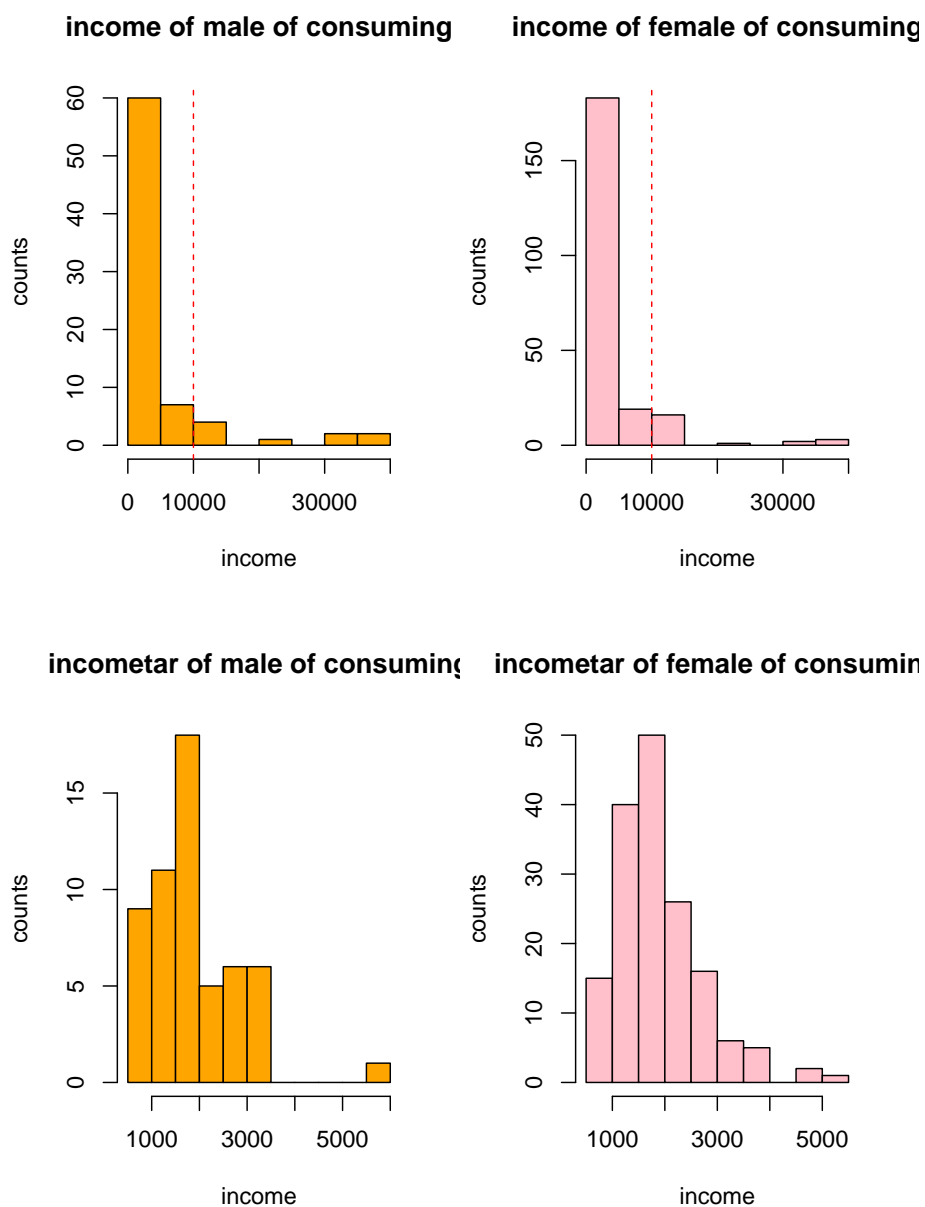
2.6.4 目标群体与为谁购买的关系



结论 7.5. 我们由图像可以看出，该群体在为谁而买方面更加侧重于为自己购买，这从某种程度上体现出了该群体以自我为中心的观念，也表现出该群体目前还不需要对家庭承担起过重的责任，只需要满足自身需求即可

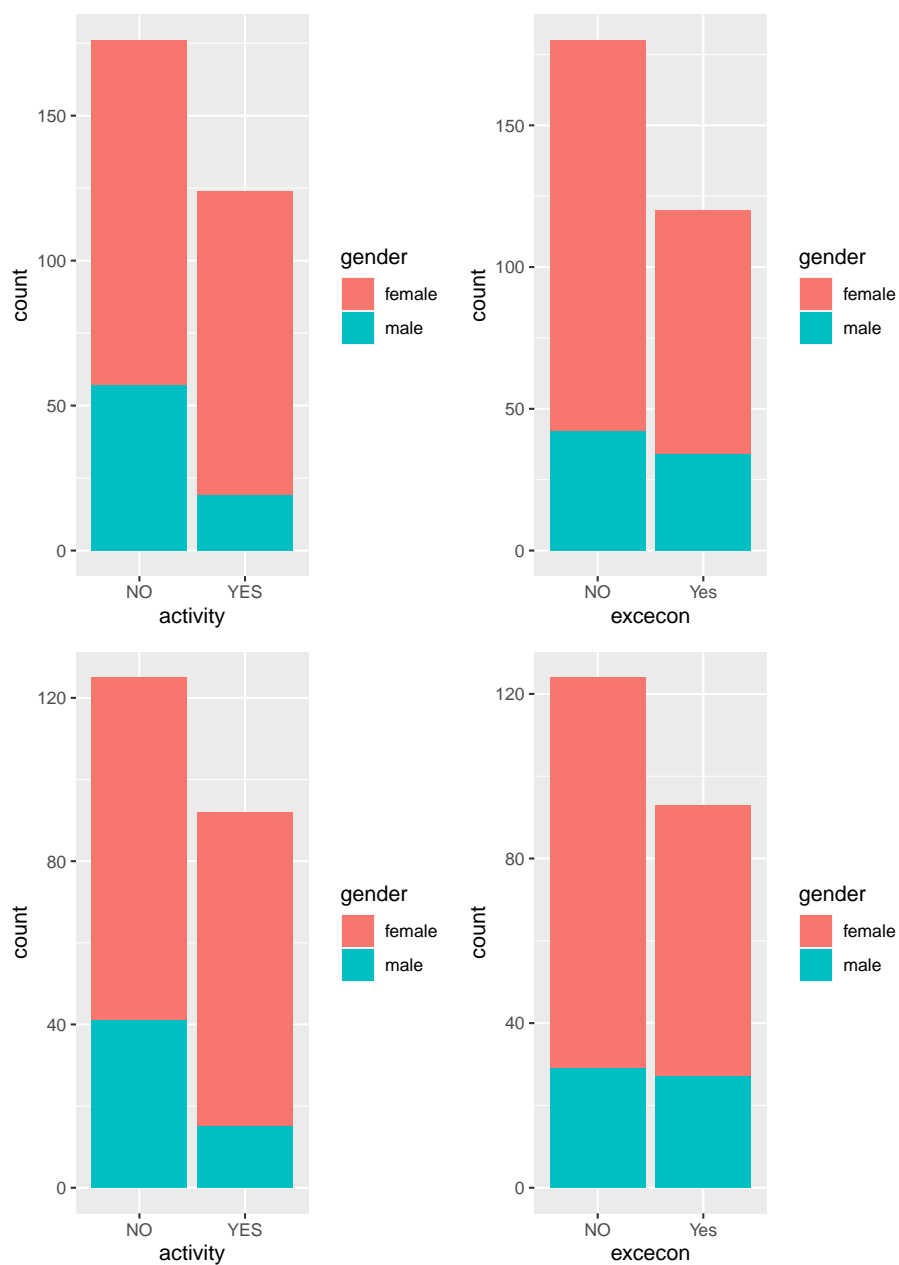
2.7 总体性别与目标群体性别 (以下简称第二类讨论群体)

2.7.1 初步分析



结论 8.1. 全体性别比不均衡，且由男女消费与收入的关系可以看出：女性的收入水平与消费的正相关性更强，由此可看出女性在面对消费时会更好的根据自己的收入安排支出，理财头脑更强. 目标群体性别比相较于总体性别比有所改善，且由男女消费与收入的关系可以看出：女性的收入水平与消费的正相关性更强，由此可看出女性在面对消费时会更好的根据自己的收入安排支出，理财头脑更强

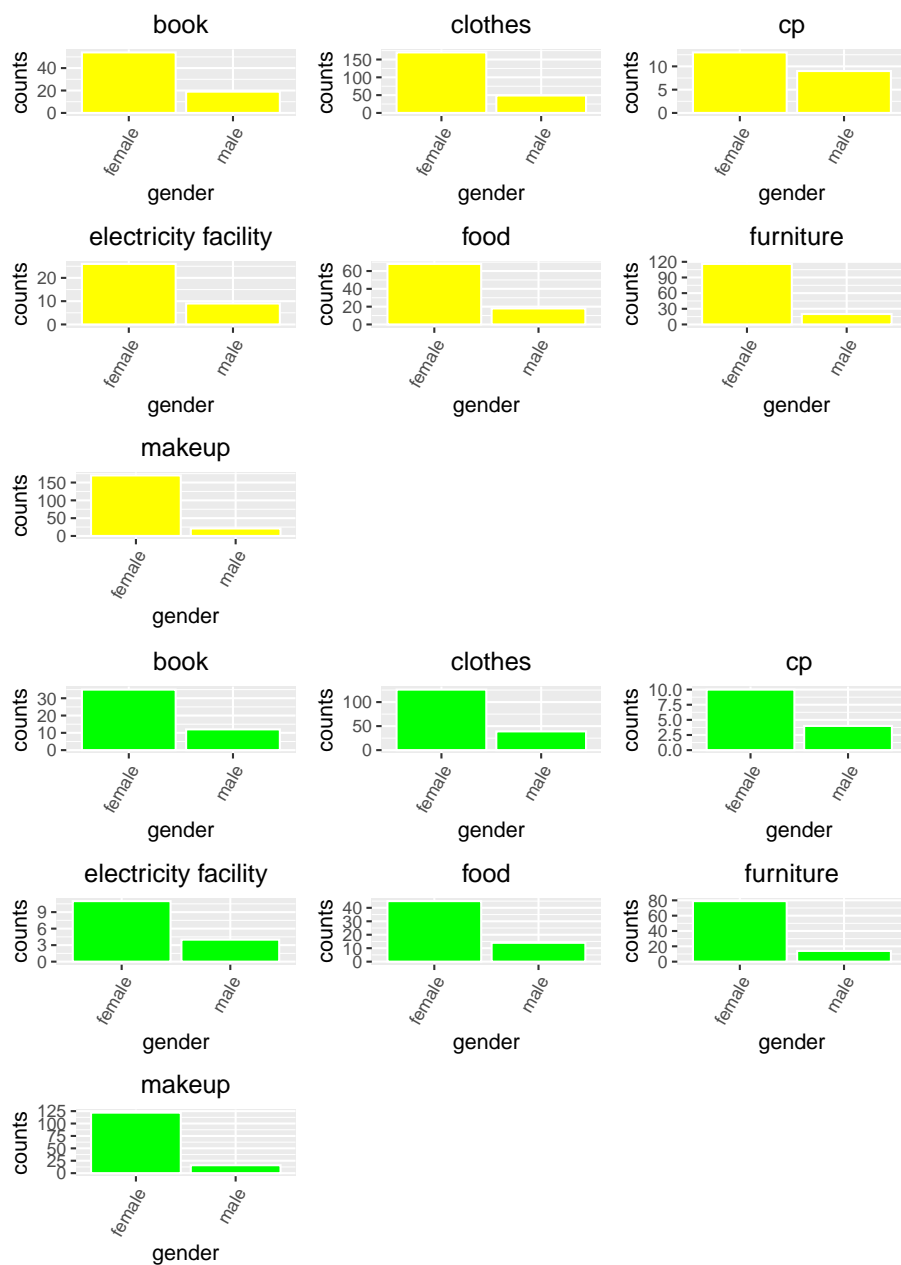
2.7.2 第二类讨论群体与是否参与活动与是否使用花呗 & 白条的关系



结论 8.2. 由上述两组图我们可以发现全部群体与目标人群在是否参与活动方面的性别特点大致相同，即男性更倾向于在购物中不参与活动，而

女性参与活动与不参与活动的比例则大致相同，这从某方面可以体现出女性在消费中对活动的倾向更强烈，而男性则不然，但对于使用花呗和白条的情况则恰好相反，这说明群体使用花呗 & 白条的影响要大于参与活动的影响

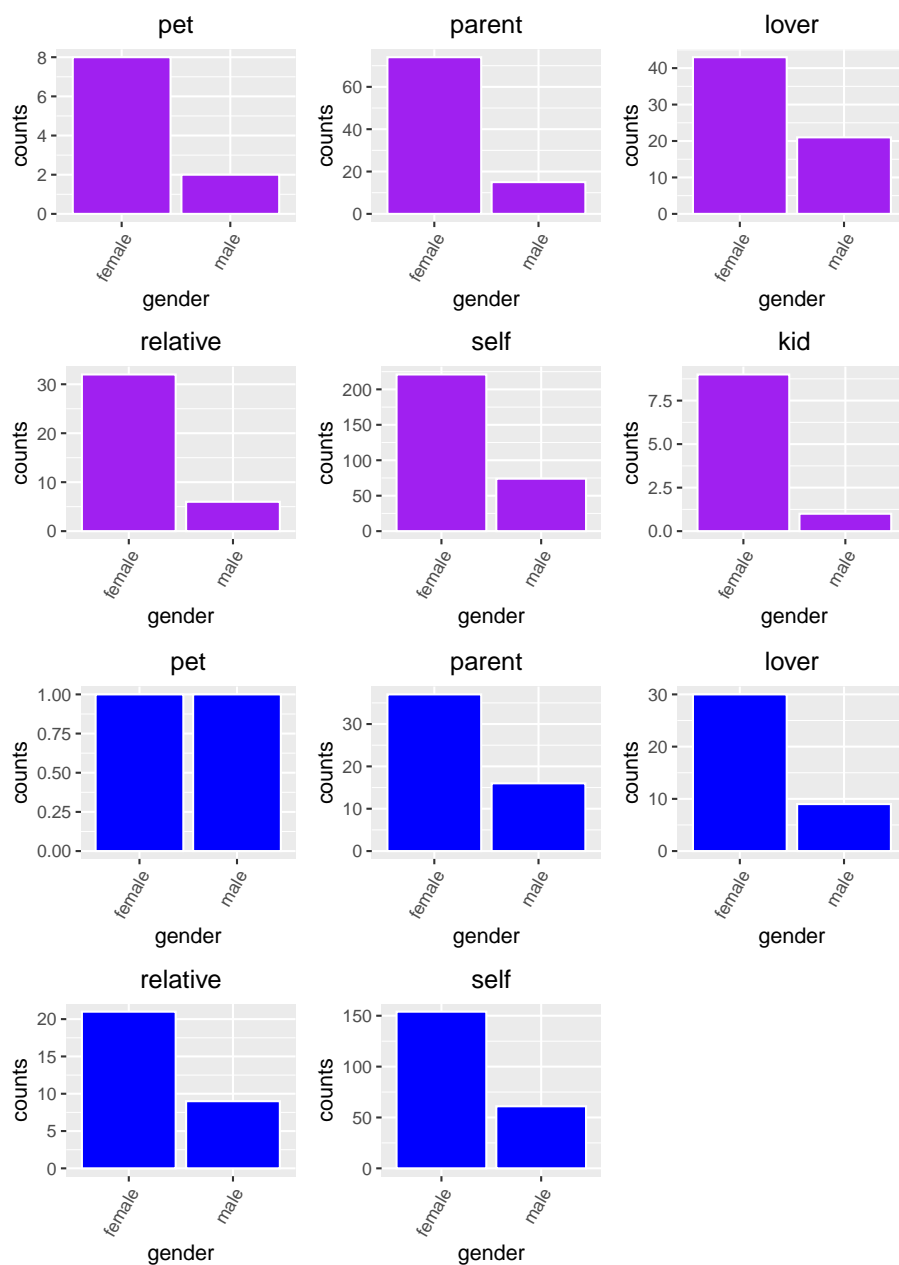
2.7.3 第二类讨论群体与所购商品种类的关系



结论 8.3.

- 对全体人群：我们可以由图像看出，对任意种类，女性的购买力均大于男性，这一方面说明了双十一参与购物的女性普遍偏多，另一方面该数据的产生也与样本量本身的性别偏颇有关，同时我们还可以发现电子产品的是唯一男性可以与女性购买量相差较小的产品，所以相对于其他更偏家具类的产品而言，男性对电子产品的偏好对于女性会更大一些，对于该种产品的了解度也会相对更深.
- 由于目标群体的男女比相对于全部样本有所改善，因而我们的分析更为准确；我们可以从图像中发现该群体任意种类商品的购买力均集中在女性群体中，男性群体的购买力呈现出极大的劣势，而且我们也可以看出目标人群女性对于服饰和化妆品的购买力仍占主导地位

2.7.4 第二类讨论群体与为谁而买的关系



结论 8.4.

- 对全体人群：从数据中我们可以看出，女性对任意人群的购买量

都远远大于男性，而男性更侧重于给恋人和自己进行购买，这从侧面反映出女性相对于男性更为顾家，而且或许还可以说明男性在恋爱中和婚姻中的重视程度不同

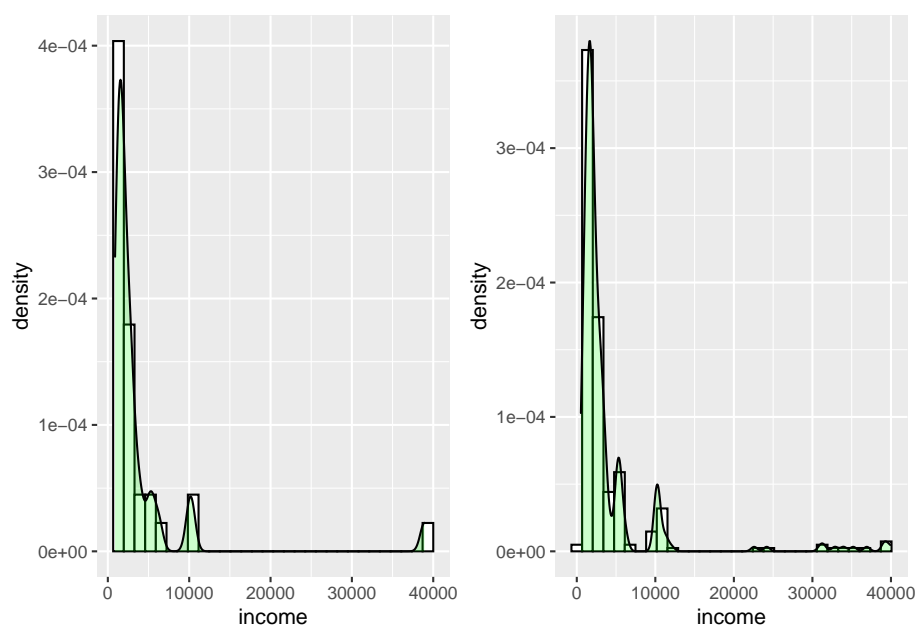
- 首先我们可以发现目标群体没有为孩子购买，这符合我们目标群体的年龄及职业特点，其次我们不难发现，和全部样本的分布相同，女性为其他人群的购买量始终大于男性群体，但就宠物而言，男性与女性的购买量一致是因为目标人群中均只有一位男性与一位女性为宠物购买了商品，对比于全部样本的分析，我们发现目标人群或许在对宠物的喜爱程度上没有太大差异，相较于全部样本，男性随着年龄递增，可能会丧失一部分对宠物的喜爱，以及宠物方面的事情已经交由家庭中女性打理的猜测

3 双十一未参与购物人群数据分析

下面我们对未参与购物人群进行分析

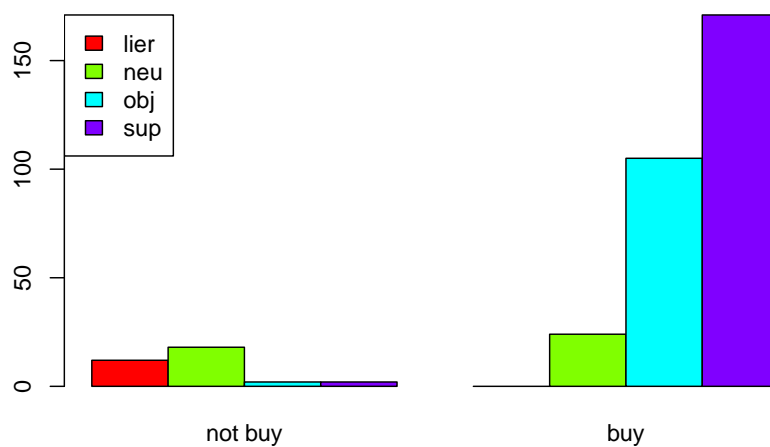
3.1 由收入情况分析

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



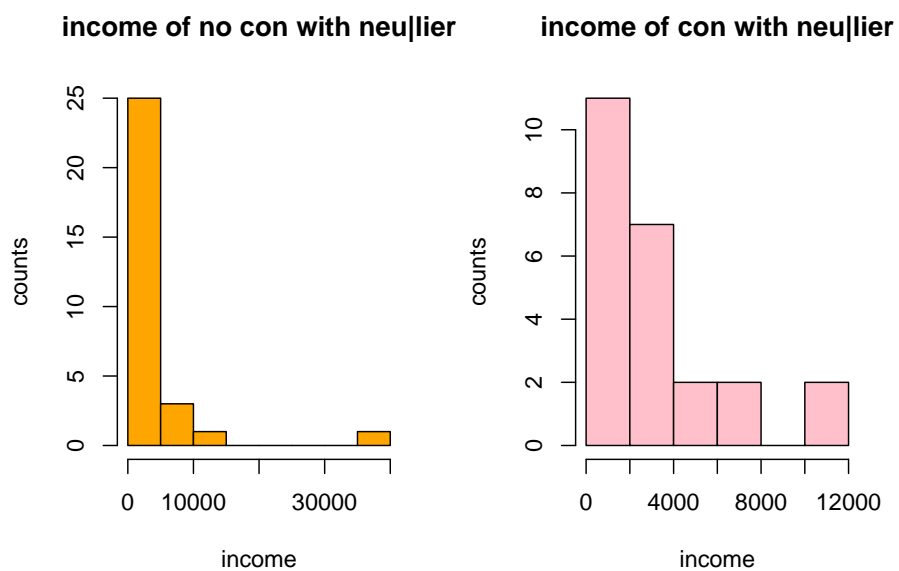
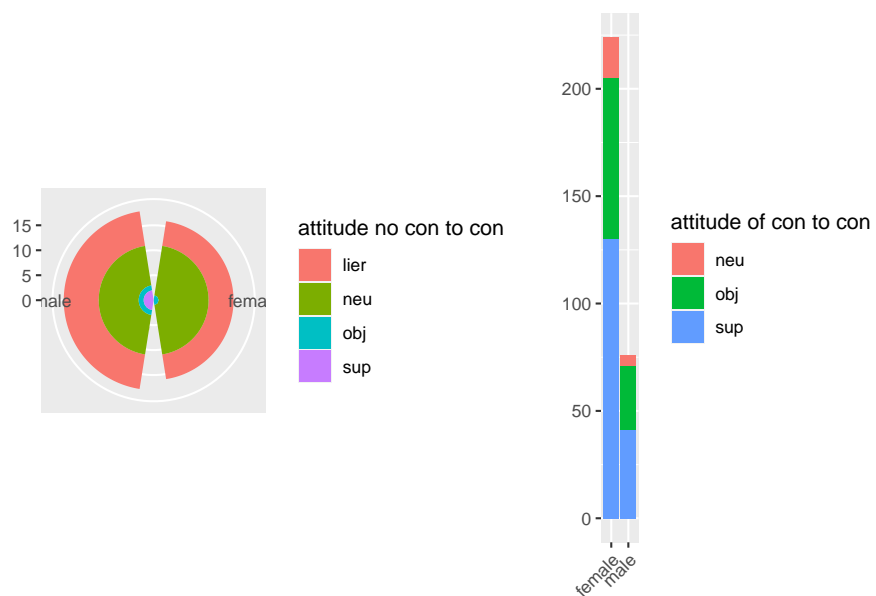
结论 9.1. 首先我们通过 `glimpse` 这个数据 (具体数据见附加代码) 能够发现从 3/4 分位数开始购物的人的收入水平要更高, 但是两者的低收入到中位数数据基本近似, 因此我们猜测不消费的人群收入要更紧凑一些, 而购物人群收入会出现较多离群值 (超高收入). 其次在画图过程中我们发现购物人群有多个离群点, 说明高消费人群中购物的比较多, 同时由箱子的区间 (蓝色线) 可以看出不购物人群的低收入人群要比购物人群的低收入人群收入要低

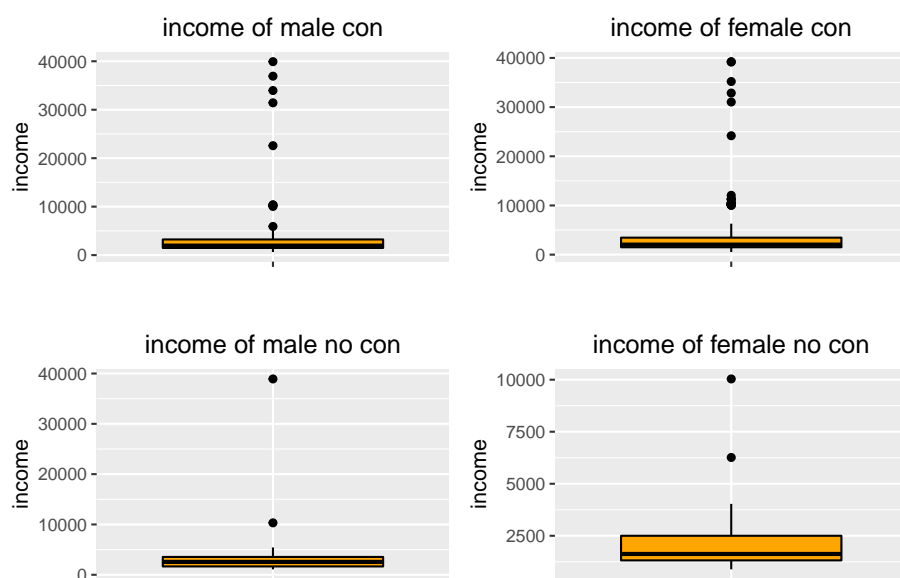
3.2 由评价分析



结论 9.2. 不购物的人并不是十分反对双十一剁手，而是对于购物比较理性，甚至于说漠不关心（说谎的人可能是填表写错了，连自己购物与否都不记得），不购物对于双十一支持和反对的人数相比于购物的人具有很大差距

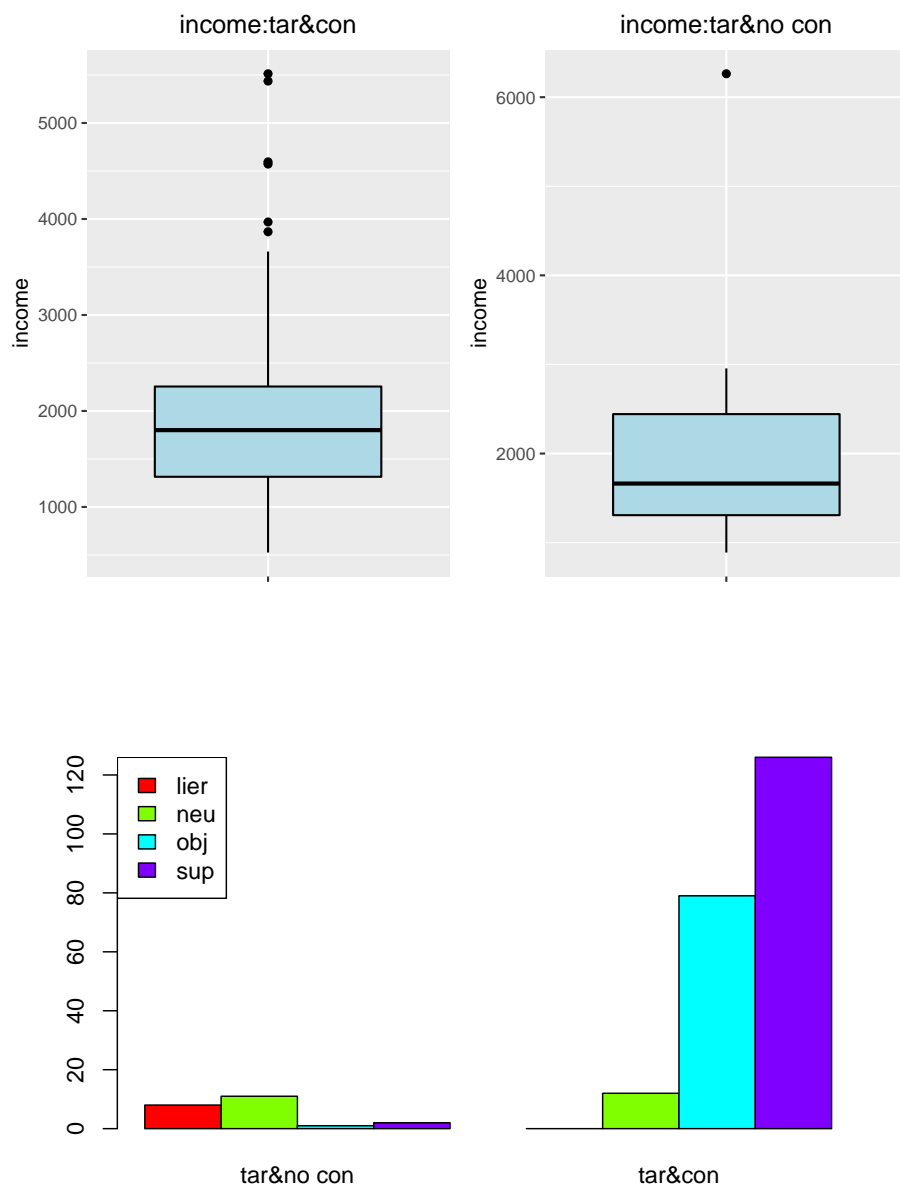
3.3 由男女比例 ~ 消费评价分析

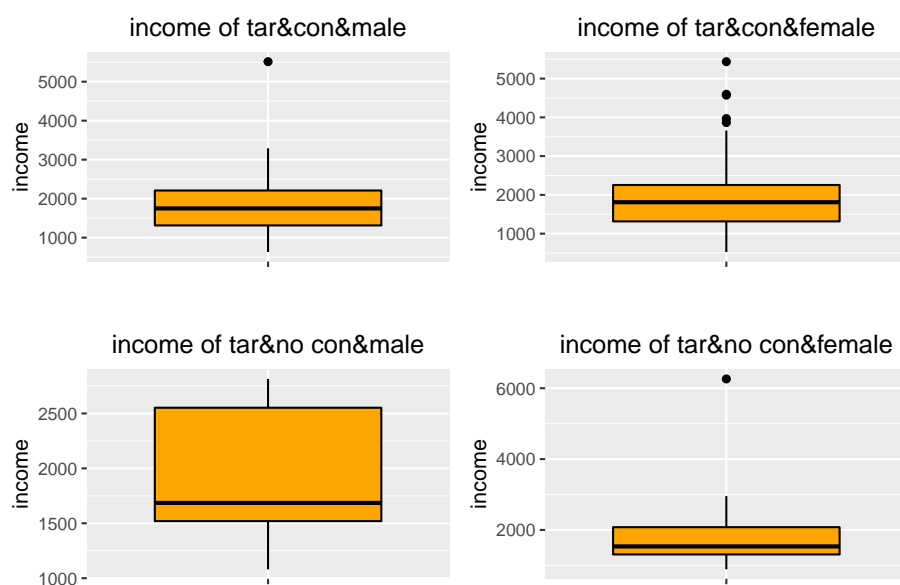




结论 9.3. 我们首先看到不消费中的女性支持人数比例明显高于男性支持比例，其次在这部分分析中所取男性样本尽管不参与消费，但是仍有小部分对于双十一消费是支持的，相比于男性，女性有更多的一部分对双十一是持有中立或者反对态度的，女性更加“淡定”，进一步我们看到不消费的女性收入较低，并且女性更理智，而且我们通过分析消费和不消费两类人中评价和收入的关系，我们可以发现不剁手且收入低的人群的评价会更加中立，而剁手的人群的中立评价对于收入则呈现出较为分散的状态，这说明收入对不剁手群体的评价的影响会更大一些，而对于剁手人群则影响较弱，(箱型图请参照本文件第一幅箱型图，对比来看) 我们还可以发现消费总体（即不考虑 18-25 岁 & 全日制学生）的收入水平要远高于不消费总体的收入水平，值得注意的是不消费的女性收入要比男性更偏低集中一点

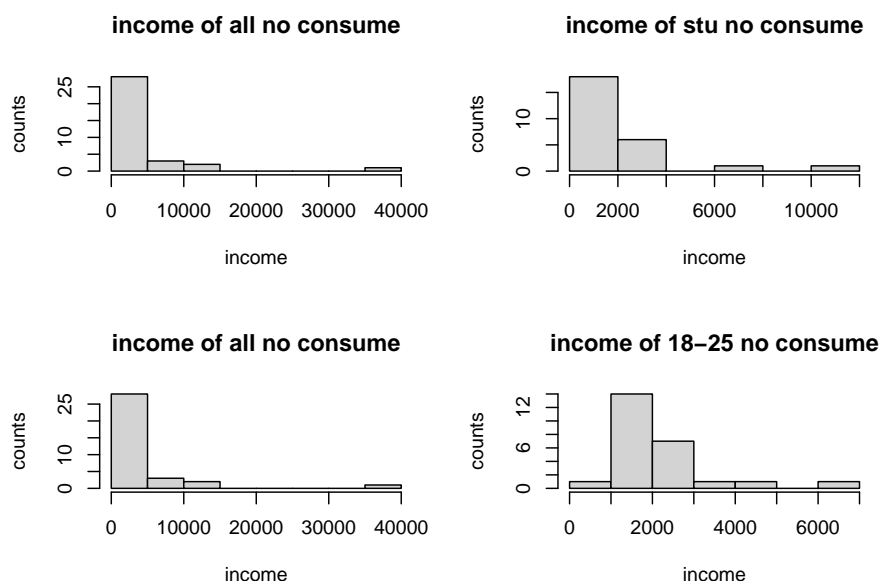
3.4 目标人群分析





结论 9.4. 我们首先可以明显看到消费目标人群的收入水平比不消费的目标人群要高，进一步我们发现消费目标群体他们收入水平高，同时热衷于给出对双十一消费的看法，支持和反对相较于中立有更多人，不消费目标群体收入较低，对双十一活动较为“漠不关心”，支持与反对的人数十分少，大多抱有中立的态度，其次因为不消费的人群中评价大多数是 neu 或者 lier（表示漠不关心）这里再分男女毫无意义，故这里只着重分析收入水平，我们得到以下结论：在第一组图中我们可以看出消费目标人群的收入较为离散，结合第二组图我们可以发现目标群体的女性收入比男性要离散，因此离散的点体现在女性群体，且结合上一步分析我们发现该处规律和上述分析中的男女规律类似，不消费女性的收入要相比于男性偏低集中，但是最后的结果是男性的相对高收入将总体的收入水平提高，但是不消费群体中不具有这样的现象

3.5 职业



结论 9.5. 我们首先可以看出学生平均收入相对要低 (该处也可从 summary 看出), 而且我们由进一步分析发现学生的态度跟总体的态度是一致的, 因而我们用学生态度代表总体可以看出学生群体平均收入还是总体偏低的

4 整体总结

总结.

1. 对于双十一消费的人群:

1. 经济条件相对较好的群体在购物时表现得更为理性, 导致其更加支持双十一购物活动。
2. 双十一购物活动带来了显著优惠, 但促销范围局限于服饰, 化妆品类, 而对大型家电及电子产品力度并不明显。大多消费者能保持相对理性, 购物的疯狂或许是商家的促销手段, 并不完全符合社会现状

3. 为谁购物受地区影响较小。随收入水平提升，消费者更愿意为孩子以外的人群选购，对于孩子，消费者对于商品是否参与活动关注度小，往往会尽量满足其需求
4. 京东和淘宝仍是目前购买力最大的平台：淘宝的受众更广泛，但其信赖度低于京东；同时在电子产品和家用电器方面，传统购物仍保持优势
5. 拼多多的拼单吸引了一批消费者，而京东和淘宝的大量消费者表明二者已经拥有自己的消费群体。尽管使用花呗白条及其他活动一定程度上影响平台收入，但是以上二者受众数量并未产生较大起伏，有利于平台的发展。
6. 18-25 岁全日制学生大部分由于收入限制能够理性消费。但仍有一部分受平台是否允许透支型消费影响大于平台活动，超前消费在该部分人群中尤为突出。目标群体与总体无较大差异
7. 女性相比于男性消费更加理智，同时购买力更强，只有电子产品二者购买数量均衡；在购买产品方面，男性较偏爱于为恋人购物，且随着时间增长，会丧失对宠物的喜爱（推测会交由女性打理），而女性对各种品类均有涉猎。最后女性对活动参与度高于男性而男性对透支型软件使用更加频繁

2. 两种人群的对比：

1. 不消费人群的收入分布较为密集，但收入水平低于消费人群，不购物人群的最低收入要低于购物人群低消费
2. 不购物人群大多对双十一购物持中立态度，而非大部分反对
3. 持中立态度者购物群体收入较为离散，而不购物者大多收入较低
4. 不购物者持支持态度的女性较多，这与购物人群的规律是一致的，可以看出女性对于购物无论是否消费仍心存希望
5. 购物的目标群体（18-25 岁全日制学生）对于双十一更乐意发表自身的看法，而不购物者大多持中立态度（或漠不关心）
6. 消费群体中的收入，男女相差不多，但是得到的总体收入数据比较离散，说明目标群体外的其他人群收入更加离散；不消费的群体中男性收入要比女性高，得到的总体收入较为集中，这是由于男性样本数量较少造成的

5 双十一购物推荐

- 平台选择: 对于家电和电子产品等更推荐选择京东及线下购物场所, 对于服饰及化妆品类产品, 更推荐选择淘宝平台
- 预期消费设置: 对于需要采购家电等大型电器的消费者而言, 需要做好消费水平高的心理准备, 以防止实际消费远超于预期消费
- 购物种类: 对于服饰化妆品类产品, 更推荐参与双十一的选购, 而对于家电类产品则无需等待双十一的购物时机

6 作业分工

尚尔淦同学主要负责数据清洗, 报告的录入, 代码方面主要负责绘制预期消费 & 实际消费, 为谁购物, 购物平台相应的与其他因素的关系, 以及未购物人群的分析

彭小雨同学主要负责报告以及结论的撰写, 代码方面主要负责购物种类, 年龄 & 性别 & 职业相应的与其他因素的关系, 以及省市区域的分布讨论, 同时也负责部分数据清洗