**发票中体现的企业经营情况**

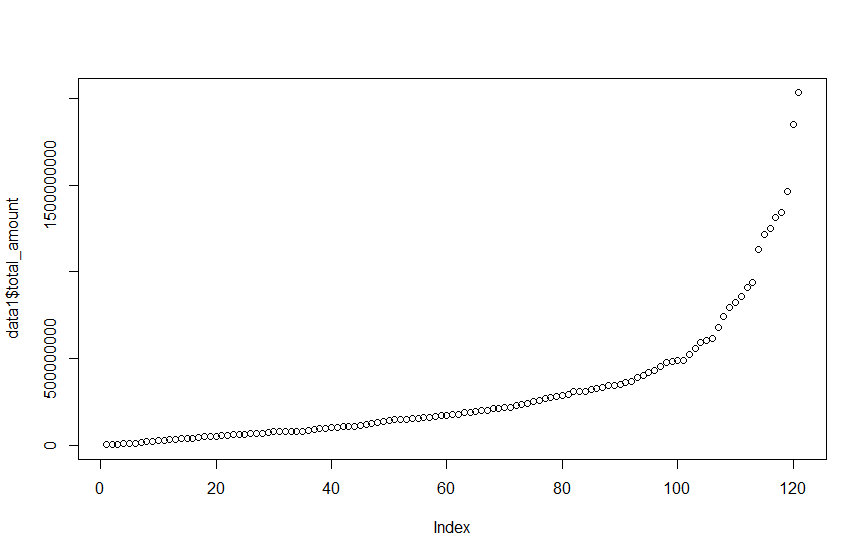
1. 介绍

此报告的目的是希望从发票数据当中挖掘有价值的信息，以用于风控的决策。所有数据来源于Shengye\_invoice里的rtc\_invoice, customer, rtc\_invoice\_detail\_list三张表中。前面的1，2点为基本的统计性图标，用于总览数据的分布，提供不了风控上的实际意义。3，4，5，6点为根据已有的数据发掘出可用于决策的指标。主要使用SQL和R语言的package，代码已附在最后的附件中。

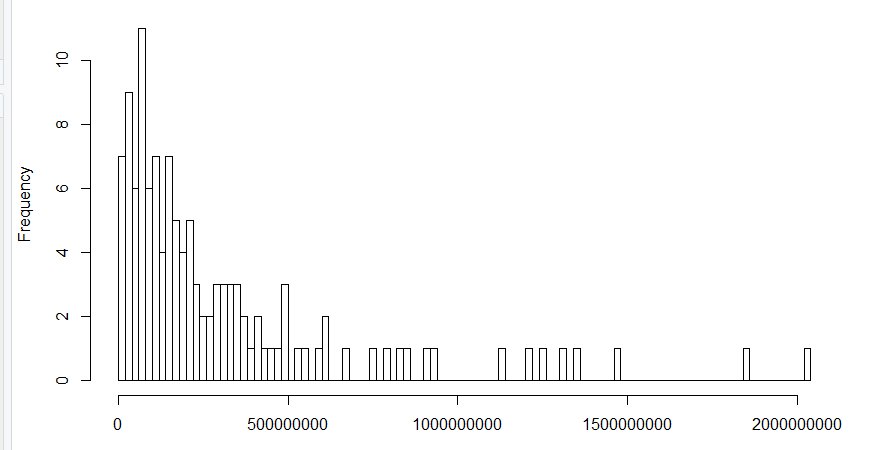
**1.**

**每个公司一年的发票总金额:**

1.1各公司发票总额散点图



1.2公司销量（总开票额）带频数的分布区间



|  |  |
| --- | --- |
| **2.各公司全年发票总额**  销售额 | 销售方 |
| 539018.45 | 成都众腾川渝供应链管理有限公司 |
| 5838583.46 | 成都中富伟业机械设备租赁有限公司 |
| 5974431.45 | 北京信远欣得酒店管理有限公司 |
| 8013448.39 | 广州市瑞智系统集成有限公司 |
| 10442708.15 | 江苏拓野供应链管理有限公司 |
| 14710170.54 | 重庆瀚盛拓实业有限公司 |
| 20500221.73 | 贵州金益捷洗涤有限公司 |
| 20685686.13 | 盛鹏商业保理有限公司 |
| 27666076.15 | 盛业（深圳）商业保理有限公司 |
| 28358934.68 | 中建三泰（北京）国际贸易有限公司 |
| 29857132.13 | 上海衍翊国际货物运输代理有限公司 |
| 34659825.24 | 贵州凯方盛商贸有限公司 |
| 35488846.44 | 广州兴合贸易有限公司 |
| 36533193.54 | 北京汇德信合商贸有限公司 |
| 37182657.94 | 山东中通宝业经贸有限公司 |
| 45628418.34 | 上海苒信实业有限公司 |
| 49687477.72 | 广州市九星建材贸易发展有限公司 |
| 50049389.84 | 山西九州通医药有限公司大同分公司 |
| 51795852.31 | 北京中京建业投资发展有限公司 |
| 54491786.47 | 山西九州通医药有限公司忻州分公司 |
| 56192428.21 | 深圳市中福科技发展有限公司 |
| 60791803.69 | 长沙恒万建材贸易有限公司 |
| 61110211.69 | 重庆艺腾物资有限公司 |
| 61217315.56 | 武汉市鑫博源物资有限责任公司 |
| 65179863.21 | 北京京北永利贸易有限公司 |
| 67107155.44 | 嘉华创盛国际贸易（北京）有限公司 |
| 67416391.92 | 重庆欧古商贸有限公司 |
| 71883981.05 | 深圳市泰恒源贸易有限公司 |
| 75791197.52 | 重庆金铜宏钢物资有限公司 |
| 75974265.78 | 广东博望实业有限公司 |
| 77438471.24 | 天津津能伟业科技发展有限公司 |
| 77480278.43 | 重庆普湄兴商贸有限公司 |
| 80076403.12 | 湖南盛世嘉华商贸有限公司 |
| 80770251.38 | 深圳市天瑞兴业投资发展有限公司 |
| 83366836.69 | 深圳市汇佰裕贸易有限公司 |
| 89616772.39 | 北京京唐联合钢铁物资有限公司 |
| 94811449.12 | 北京嘉启鸿商贸有限公司 |
| 94819351.84 | 盛业信息科技服务（深圳）有限公司 |

**2.1各公司平均发票金额**

各公司平均发票金额 销售方

"668306.264525" "上海大展金属材料有限公司"

"82510.702242" "上海苒信实业有限公司"

"113289.478009" "上海诚展实业有限公司"

"460115.649176" "上海语骏实业有限公司"

"89179.039874" "中建三泰（北京）国际贸易有限公司"

"386414.790714" "五矿联合（北京）国际贸易有限公司"

"629927.934406" "北京东方明贵贸易有限公司"

"92658.054222" "北京中京建业投资发展有限公司"

"573954.449133" "北京中科辉腾科技有限公司"

"102175.246344" "北京中钢新兴投资发展有限公司"

"92406.030461" "北京九联伟业投资发展有限公司"

"90613.521122" "北京京唐联合钢铁物资有限公司"

"2767.221607" "北京信远欣得酒店管理有限公司"

"797366.496887" "北京力晟通宇商贸有限公司"

"610384.822343" "北京华中冶贸易有限公司"

"813834.277617" "北京华泰明珠贸易有限公司"

"76565.711708" "北京吉兆年商贸有限公司"

"787205.186725" "北京同兴德贸易有限公司"

"88941.321876" "北京嘉启鸿商贸有限公司"

"672971.476330" "北京天铖德瑞贸易有限公司"

"764896.694446" "北京市中联基业贸易有限公司"

"70895.708778" "北京德润华贸科技有限公司"

"57351.952182" "北京汇德信合商贸有限公司"

"91134.921408" "北京沧海联业科技有限公司"

"503901.480091" "北京澳隆德科技发展有限公司"

"729051.915673" "北京金世佳铭商贸有限公司"

"454165.051918" "北京鑫易润科技有限公司"

"820865.756047" "北京隆昌恒鑫商贸有限公司"

"107074.097303" "北京顺通力创商贸有限公司"

"91801.854227" "嘉华创盛国际贸易（北京）有限公司"

"83278.053141" "嘉华创盛国际贸易（福建）有限公司"

"437559.734897" "天津东方创展商贸有限公司"

"96416.325458" "天津市亿旺钢铁贸易有限公司"

"89524.244208" "天津津能伟业科技发展有限公司"

"87193.352962" "天津肯润钢铁贸易有限公司"

"38070.369298" "安徽乐嘉医药科技有限公司"

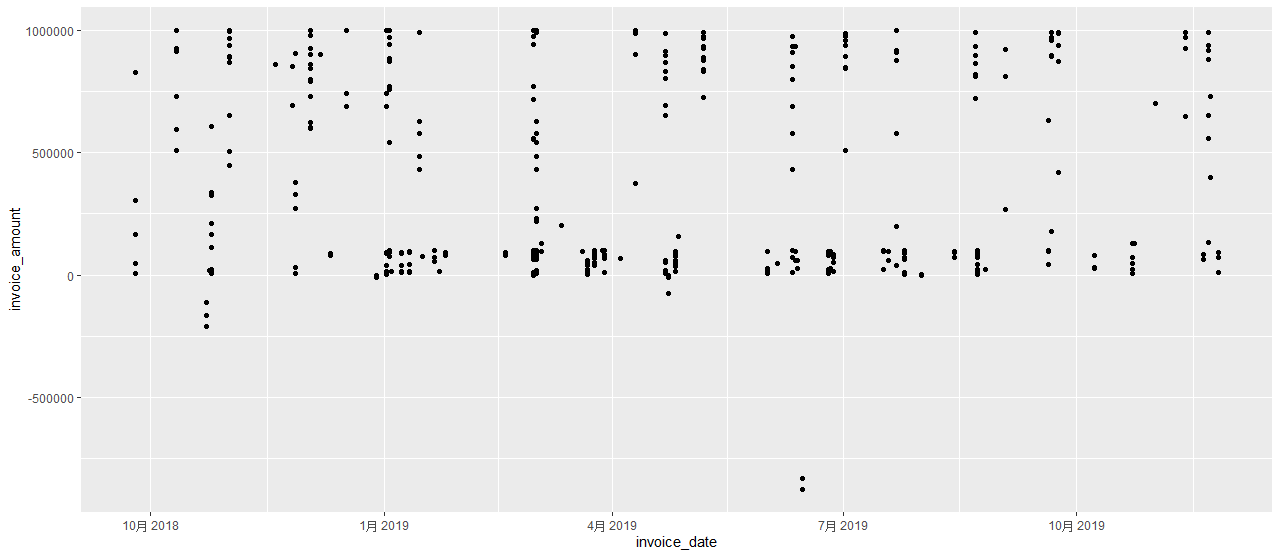
"88319.852589" "山东中通宝业经贸有限公司"

"928380.639856" "山东隆腾盛业商贸有限公司"

"781650.671992" "山西源晟发工贸有限公司"

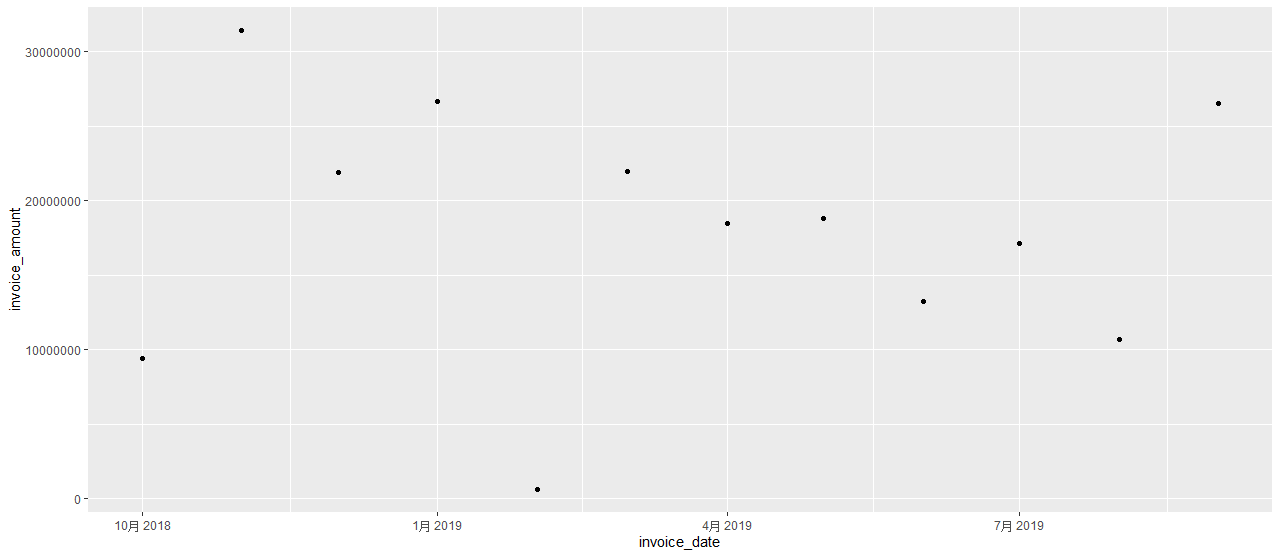
**3.1艾尔斯建筑材料（深圳）有限公司发票额变动情况**

在此选取一家有一定数据量的公司用于示例，之后可利用脚本输出所有公司的这些指标情况。



高额发票集中固定于每个月，低额发票时间相对零散，总体上说开具发票时间点具有一定规律性，但是需要多年数据进行同比能够得到一些开票量时间点上的反馈。这张图在可视化上还可以进行优化，X轴标刻度无法容易的看出是该月的哪个时期(上中下旬)。

**3.2艾尔斯建筑材料（深圳）有限公司发票额变动情况（按月）**



将发票额度按月集中体现更能够反映出该公司开票额上的季节性。从该图可以看出2月，10月，8月为该公司业务上的淡季。年末是一个业务上的高潮。与上一张一样，需要多年的数据纵向对比，或者将其与相同类型的公司进行横向比较。

**4 发票异常比例:**

|  |  |  |  |
| --- | --- | --- | --- |
| 异常发票数 | 总发票数 | 异常发票比例 | 公司名 |
| "68" | "3269" | "0.020377584" | "重药宁夏" |
| "17" | "2000" | "0.008428358" | "重药河南" |
| "0" | "1509" | "0" | "重庆医药集团（宁夏）有限公司" |
| "28" | "1344" | "0.020408163" | "重庆医药集团（宁夏）有限公司" |
| "41" | "1089" | "0.036283185" | "重药宁夏" |
| "28" | "1000" | "0.027237354" | "重庆医药集团（宁夏）有限公司" |
| "14" | "886" | "0.015555555" | "重药河南" |
| "0" | "869" | "0" | "上海润达医疗科技股份有限公司" |
| "0" | "801" | "0" | "重庆医药集团（宁夏）有限公司" |
| "14" | "695" | "0.019746121" | "重庆医药集团河南有限公司" |
| "0" | "650" | "0" | "上海润达" |
| "5" | "628" | "0.007898894" | "国药西藏医药有限公司" |
| "0" | "614" | "0" | "上海润达医疗科技股份有限公司" |
| "2" | "560" | "0.003558718" | "重庆医药集团河南有限公司" |
| "0" | "490" | "0" | "国药西藏医药有限公司" |
| "0" | "488" | "0" | "国药西藏医药有限公司" |
| "0" | "427" | "0" | "国药西藏医药有限公司" |
| "31" | "355" | "0.08031088" | "深圳博睿智慧建造科技有限公司" |
| "230" | "312" | "0.424354243" | "t1" |
| "0" | "311" | "0" | "重庆医药集团（宁夏）有限公司" |
| "0" | "304" | "0" | "北京瑞丰来科贸有限公司" |
| "0" | "286" | "0" | "天津物华新天商贸有限公司" |
| "0" | "281" | "0" | "上海润达医疗科技股份有限公司" |
| "2" | "273" | "0.007272727" | "嘉华创盛国际贸易（福建）有限公司" |
| "0" | "255" | "0" | "国药西藏医药有限公司" |
| "0" | "250" | "0" | "深圳市中荣建工程有限公司" |
| "0" | "219" | "0" | "广州汉鹏供应链有限公司" |
| "30" | "207" | "0.126582278" | "北京九联伟业投资发展有限公司" |

由于冲销掉的发票是不包含在发票异常比例中的，所以异常的发票比例可以作为一个单独的指标来进行考量，异常的缘由有可能是假票，发票类型不对或者操作错误等。

由于篇幅原因只显示部分公司。

**5 发票冲销比例**

5.1冲销发票总体情况



总异常发票为20338张，总发票数为969437张。

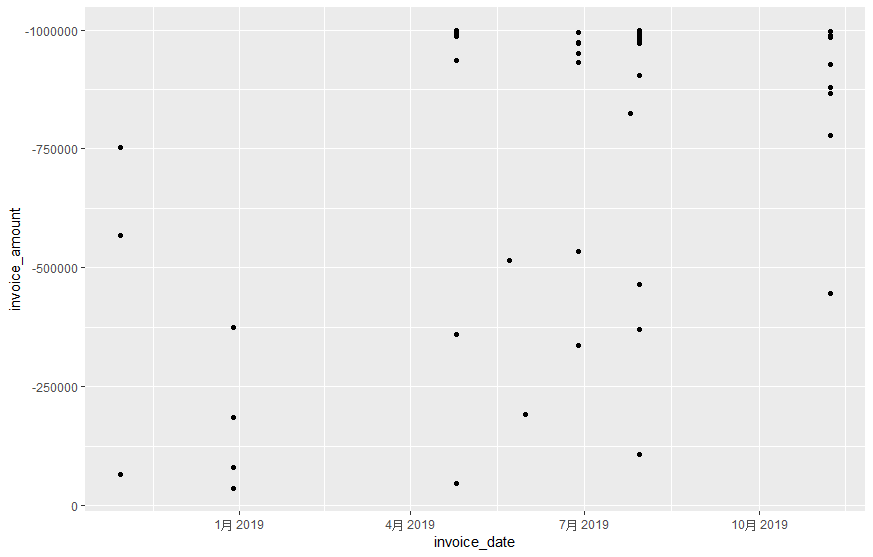
5.2 各公司异常发票

|  |  |  |  |
| --- | --- | --- | --- |
| 冲销数 | 发票数 | 冲销比例 | saler\_name |
| 0 | 1 | 0.00% |  |
| 62 | 2767 | 2.24% | 上海大展金属材料有限公司 |
| 0 | 553 | 0.00% | 上海苒信实业有限公司 |
| 39 | 7733 | 0.50% | 上海衍翊国际货物运输代理有限公司 |
| 6 | 899 | 0.67% | 上海诚展实业有限公司 |
| 10 | 461 | 2.17% | 上海语骏实业有限公司 |
| 2 | 318 | 0.63% | 中建三泰（北京）国际贸易有限公司 |
| 2 | 953 | 0.21% | 五矿联合（北京）国际贸易有限公司 |
| 2595 | 243423 | 1.07% | 内蒙古瑞康医药有限公司 |
| 51 | 2111 | 2.42% | 内蒙古诚钢物资有限公司 |
| 2 | 320 | 0.63% | 北京东方明贵贸易有限公司 |
| 0 | 559 | 0.00% | 北京中京建业投资发展有限公司 |
| 9 | 600 | 1.50% | 北京中科辉腾科技有限公司 |
| 0 | 1559 | 0.00% | 北京中钢新兴投资发展有限公司 |
| 13 | 5158 | 0.25% | 北京九联伟业投资发展有限公司 |
| 0 | 768 | 0.00% | 北京京北永利贸易有限公司 |
| 8 | 989 | 0.81% | 北京京唐联合钢铁物资有限公司 |
| 20 | 2159 | 0.93% | 北京信远欣得酒店管理有限公司 |
| 9 | 604 | 1.49% | 北京力晟通宇商贸有限公司 |
| 12 | 683 | 1.76% | 北京华中冶贸易有限公司 |
| 2 | 642 | 0.31% | 北京华泰明珠贸易有限公司 |
| 38 | 4068 | 0.93% | 北京吉兆年商贸有限公司 |
| 0 | 287 | 0.00% | 北京同兴德贸易有限公司 |
| 0 | 1066 | 0.00% | 北京嘉启鸿商贸有限公司 |
| 0 | 147 | 0.00% | 北京嘉豪鸿业商贸有限公司 |
| 0 | 722 | 0.00% | 北京天铖德瑞贸易有限公司 |
| 16 | 1120 | 1.43% | 北京市中联基业贸易有限公司 |
| 120 | 1473 | 8.15% | 北京德润华贸科技有限公司 |
| 24 | 1188 | 2.02% | 北京正和工程装备服务股份有限公司 |
| 0 | 637 | 0.00% | 北京汇德信合商贸有限公司 |
| 0 | 1904 | 0.00% | 北京沧海联业科技有限公司 |
| 10 | 1100 | 0.91% | 北京澳隆德科技发展有限公司 |
| 3 | 1285 | 0.23% | 北京金世佳铭商贸有限公司 |
| 4 | 245 | 1.63% | 北京鑫易润科技有限公司 |
| 11 | 592 | 1.86% | 北京隆昌恒鑫商贸有限公司 |
| 0 | 1435 | 0.00% | 北京顺通力创商贸有限公司 |
| 228 | 30753 | 0.74% | 吉林省天华医药有限责任公司 |
| 2 | 731 | 0.27% | 嘉华创盛国际贸易（北京）有限公司 |
| 93 | 1662 | 5.60% | 嘉华创盛国际贸易（福建）有限公司 |

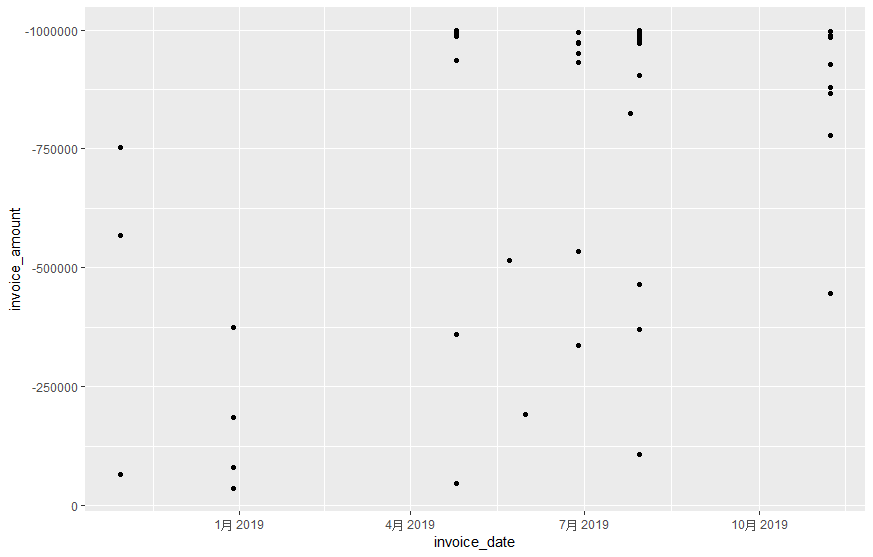
冲销发票比例也是一个很重要的指标。在临近财务重大披露的日期（如审计，尽职调查）时，公司会通过开假发票造成应收账款增多，提升自己的业务指标，并且在披露事项之后进行冲回。在此筛选掉了所有-100元以上的异常发票额度，以避免正常的失误造成的小额度发票冲销影响公司的业绩，并且在低面额发票上作假的意义不大。该筛选额度还可以调整，只是在此提示需要考虑这个细节。

由于篇幅原因只显示部分公司。

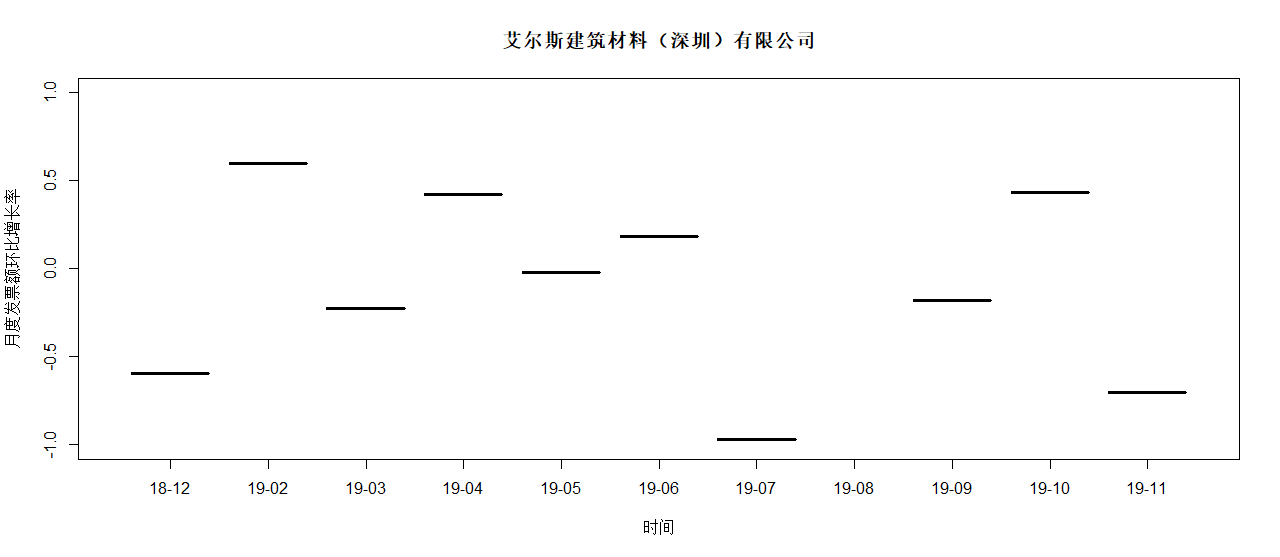
**5.3冲销金额时序变化图(上海大展金属材料有限公司)(按日)**



**5.3冲销金额时序变化图(上海大展金属材料有限公司)(按月)**



**6.发票额月度环比数据**



此图为公司开票情况变化的增长率，这是根据风控部门同事kevin的提出的一些建议制作的。能看出一些问题。

由于数据不足2年 没有办法做同比数据。

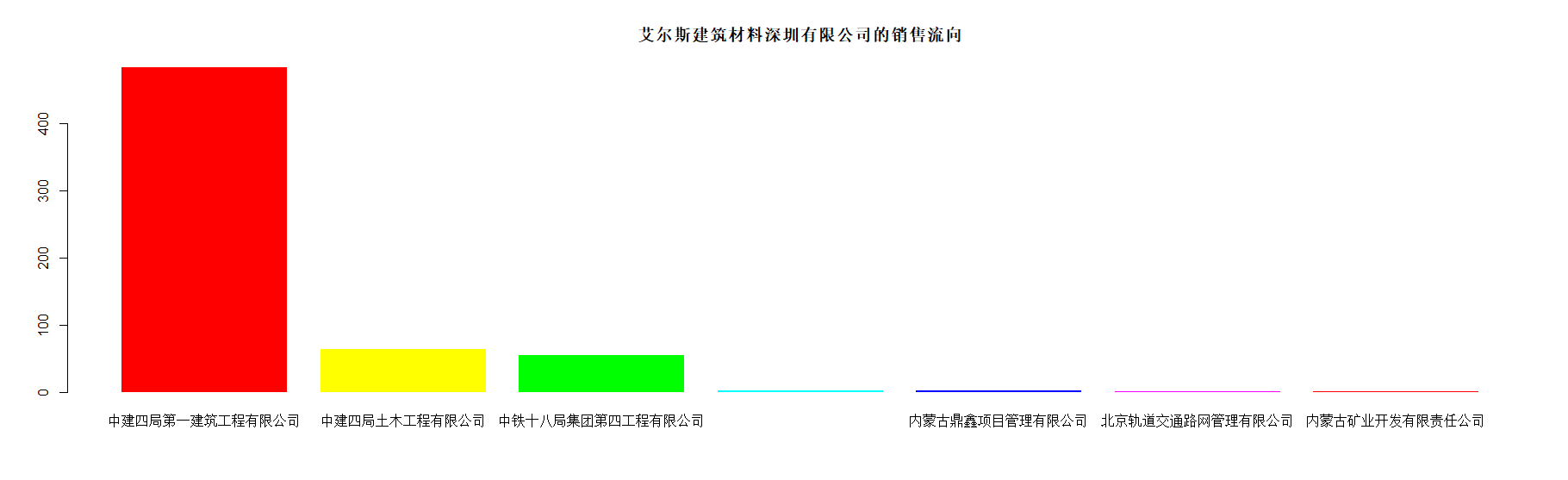
**7 各公司货物清单整理**

由于需要的数据库两表之间无法联立，暂时查询结果无法做出每家公司的所有货物数量的总结图表，还在研究。

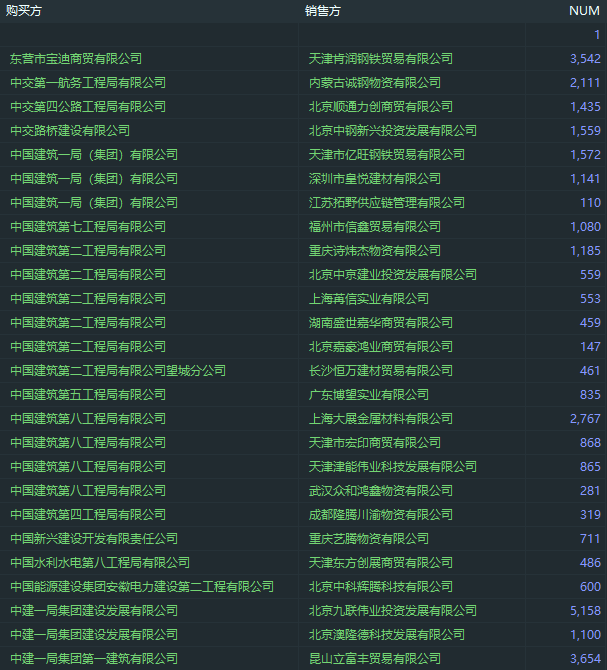
**8 销货方公司销货流向与购货方公司进货渠道(发票数量)**

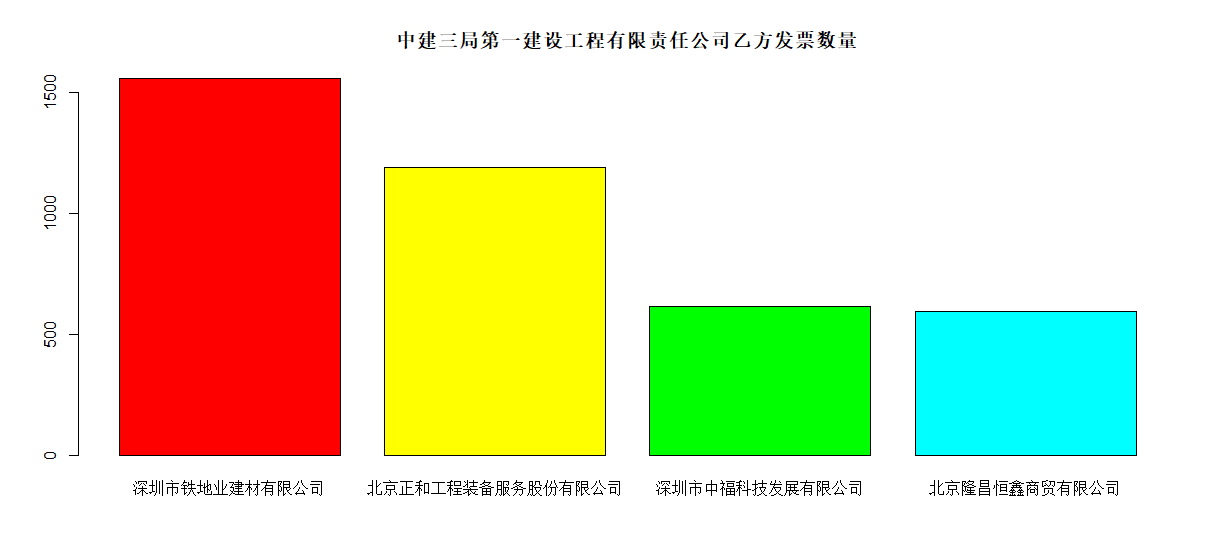
**销货方公司销货流向**





**购货方公司进货渠道**

****

****

**8.未来改善的设想**

因为发票数据是销项数据，而且开票日不等于收款日，所以我们只能看到这家公司的业务能力，而看不到这家公司的变现和短期偿债能力。如果有进项数据就可以做一些流动性和现金流上的分析，通过研究开票日与收到款项的日期可以做一些存货周转天数，流动率上的分析。

此外，基于可以监控基本户，可以尝试让基本户的流水去匹配销项发票上的数值，不仅可以得到应收账款的周转天数，也可以监控他的实际现金流的流量，实现流动性和现金流的分析。

此外，通过各个公司的发票日期和金额，只能分析出企业业务的季节性和集中度。

**附件：**

**1代码：**

SQL代码:

**SELECT** **sum**(rtc\_invoice.invoice\_amount) **AS** total\_amount,rtc\_invoice.invoice\_date,rtc\_invoice.invoice\_amount,customer.industry\_category,rtc\_invoice.saler\_name **FROM** rtc\_invoice

rtc\_invoice **LEFT** **JOIN** customer **ON** rtc\_invoice.saler\_name=customer.customer\_name

**GROUP** **BY** rtc\_invoice.saler\_name

**ORDER** **BY** rtc\_invoice.saler\_name,invoice\_date **DESC**

R代码:

hist(data1$total\_amount,breaks=122,angle=60)

hist(data1$total\_amount,breaks=122,angle=60,main = 'Taxation frequency interval')

**2代码：**

**SELECT** **sum**(rtc\_invoice.invoice\_amount) **AS** total\_amount,rtc\_invoice.invoice\_date,rtc\_invoice.invoice\_amount,customer.industry\_category,rtc\_invoice.saler\_name **FROM** rtc\_invoice

rtc\_invoice **LEFT** **JOIN** customer **ON** rtc\_invoice.saler\_name=customer.customer\_name

**GROUP** **BY** rtc\_invoice.invoice\_date

**ORDER** **BY** invoice\_date **DESC**

**2.1代码:**

SQL代码：

**SELECT** **avg**(rtc\_invoice.invoice\_amount),customer.industry\_category,rtc\_invoice.saler\_name **FROM** rtc\_invoice

rtc\_invoice **LEFT** **JOIN** customer **ON** rtc\_invoice.saler\_name=customer.customer\_name

**WHERE** customer.industry\_category=1

**GROUP** **BY** rtc\_invoice.saler\_name

**3.1代码：**

SQL代码:

**SELECT** **DATE\_FORMAT**(rtc\_invoice.invoice\_date,'%y/%m'),rtc\_invoice.invoice\_amount,customer.industry\_category,rtc\_invoice.saler\_name **FROM** rtc\_invoice

rtc\_invoice **LEFT** **JOIN** customer **ON** rtc\_invoice.saler\_name=customer.customer\_name

**WHERE** customer.industry\_category=1 **AND** rtc\_invoice.saler\_name='艾尔斯建筑材料（深圳）有限公司'

**GROUP** **BY** **YEAR**(rtc\_invoice.invoice\_date),**MONTH**(rtc\_invoice.invoice\_date)

**ORDER** **BY** rtc\_invoice.saler\_name,invoice\_date **DESC**

R代码:

ggplot(GGG,aes(x = invoice\_date,y =invoice\_amount)) +

+ geom\_point()

**3.2代码：**

SQL代码:

**SELECT** **DATE\_FORMAT**(rtc\_invoice.invoice\_date,'%y/%m/%01') **AS** invoice\_date,**sum**(rtc\_invoice.invoice\_amount) **AS** invoice\_amount,customer.industry\_category,rtc\_invoice.saler\_name **FROM** rtc\_invoice

rtc\_invoice **LEFT** **JOIN** customer **ON** rtc\_invoice.saler\_name=customer.customer\_name

**WHERE** customer.industry\_category=1 **AND** rtc\_invoice.saler\_name='艾尔斯建筑材料（深圳）有限公司'

**GROUP** **BY** **MONTH**(rtc\_invoice.invoice\_date)

**ORDER** **BY** rtc\_invoice.saler\_name,invoice\_date **DESC**

R代码:

ggplot(data4,aes(x = invoice\_date,y =invoice\_amount)) +geom\_point()

**4代码:**

SQL:

**SELECT** data\_top\_num.unusual\_quantity,data\_top\_num.normal\_quantity, (data\_top\_num.unusual\_quantity/(data\_top\_num.unusual\_quantity+data\_top\_num.normal\_quantity))+'%' **AS** percentage **FROM** data\_top\_num

**5.2代码：**

SQL代码:

**select** **sum**(**case** **when** r.invoice\_amount<-100 **then** 1 **ELSE** 0 **end**) **as** num,**count**(\*) **as** allnum ,r.saler\_name

**from** rtc\_invoice r

**GROUP** **BY** r.saler\_name;

**5.3代码:**

ggplot(TEST,aes(x = invoice\_date,y =invoice\_amount),inplace=FALSE) +ylim(max(invoice\_amount),min(invoice\_amount))+geom\_point()

**6代码：**

R代码:

text.x <- c("18-12","19-02","19-03","19-04","19-05","19-06","19-07","19-08","19-09","19-10","19-11")

t<-exp(diff(log(invoice\_amount)))-1

DF<-data.frame(text.x,t)

A<-plot(DF,ylim=c(-1,1),type='p',xlab="时间",ylab="月度发票额环比增长率",Col="RED",main="艾尔斯建筑材料（深圳）有限公司")

8代码:

SQL代码:

**SELECT** rtc\_invoice.saler\_name, rtc\_invoice.buyer\_name,**COUNT**(\*) **AS** NUM **FROM** rtc\_invoice

**GROUP** **BY** rtc\_invoice.buyer\_name,rtc\_invoice.buyer\_name

**ORDER** **BY** rtc\_invoice.saler\_name

R代码:

group1 <- split(ttt,ttt$saler\_name)

group2<-as.data.frame(group[1])

barplot(group2$'艾尔斯建筑材料.深圳.有限公司.NUM',names.arg = group2$'艾尔斯建筑材料.深圳.有限公司.buyer\_name',col=rainbow(6),border = NA,main='艾尔斯建筑材料深圳有限公司的销售流向')