从金融市场到区块链

R语言商业实践

青萌数海CTO – 张丹

第11届中国R大会(上海)

目录

CONTENTS

01 思维模式

02 金融市场

03 区块链

PART

01

思维模式

跨界，是需求要综合素质的长期积累。

认知

- R语言适合做什么，不适合做什么？
- 先选择场景，还是先选择算法？
- 得到了什么结果，解决什么问题，价值如何体现？

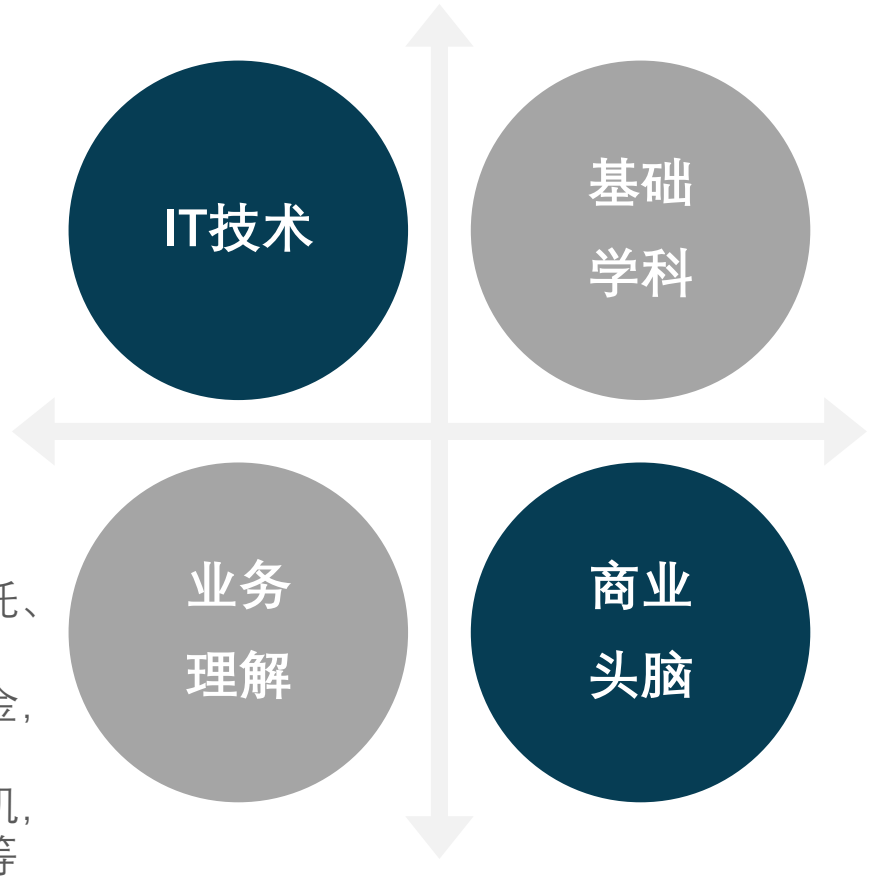
复合型知识体系

IT技术

编程能力，操作系统，网络，大数据技术，挖掘算法，爬虫，数据库，内存控制，高性能计算，软件使用等

业务理解

金融机构：银行，证券，保险，信托、基金等
金融市场：股票、债券，期货，基金，外汇，P2P，理财、利率、汇率等
区块链：交易所，公链，矿场，矿机，项目方，媒体，数据，基金、散户等



基础学科

初等数学，高等数学，线性代数，离散数学，概率论，统计学，计量经济学、投资学、金融学

商业头脑

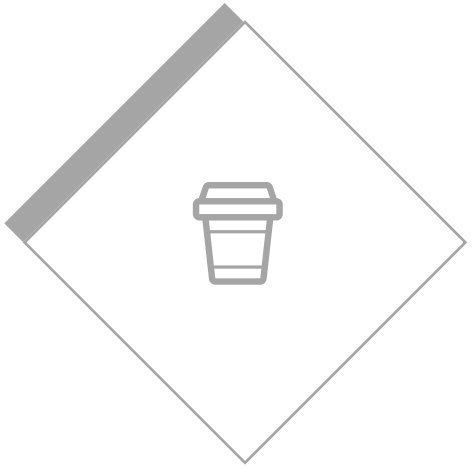
商业模式，价值定义，价值探索，价值挖掘，价值实现

如果你和我一样，欢迎加入我的团队!

解决方案

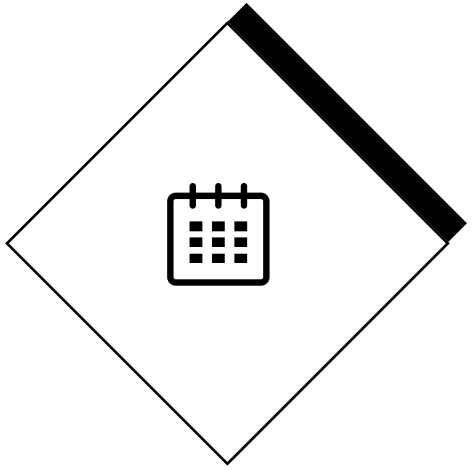
单一维度业务数据

登记数据，GDP，利率，日活，月活，销售额，销售量，价格



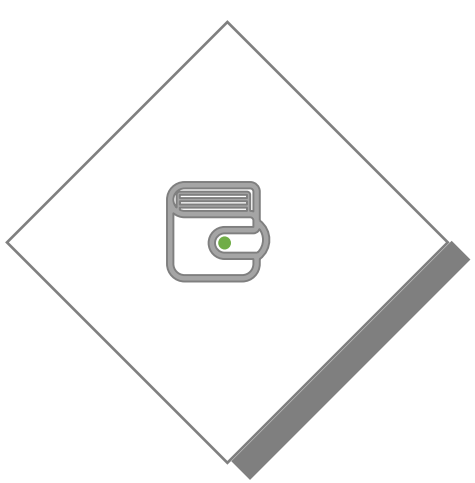
多维度复杂业务数据

统计套利，投资组合，反洗钱



单一维度大数据

出租车运行轨迹，电商商品销售，新闻推荐，好友关系



多维度大数据

如何公共交通运行效率，全球市场货物贸易，提升Facebook广告收入

职业方向

INVESTMENT RETURN

系统架构师 VS 数据科学家

发现价值，比实现需求更重要！

PART

02

金融市场

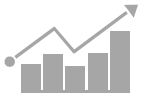
市场在变化，要跟上市场的变化

金融市场基础理论

资本资产定价模型 CAPM

alpha, beta, 资本市场线，最优投资组合假设，大类资产配置

套利定价理论 APT



配对交易，跨市场交易，统计套利

多因子方法论

选股票，选基金，因子分类，因子有效性检查

投资组合优化 black-litterman

贝叶斯分析，主观客观结合，条件概率，隐含收益率

投资组合

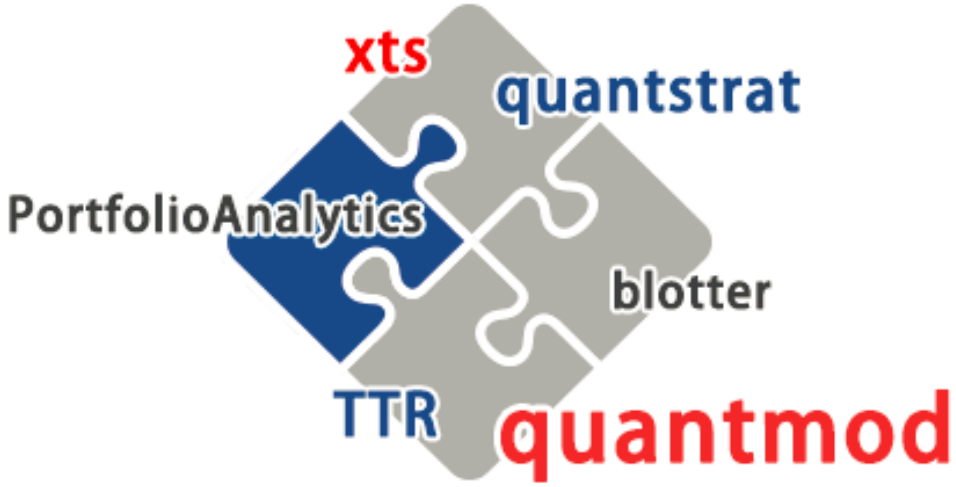
近百个点来说，假设每次要配置5种资产做投资组合，那么就是75287520种组合方法；如果配置10种资产，可选方案就是一个非常大的数字了。

x轴为收益率的标准差表示风险，y轴为收益率。

图中的点构建了可投资区域，每个点代表一个可投资产品，每条虚线连接的点的集合，就是一个有效的投资组合。

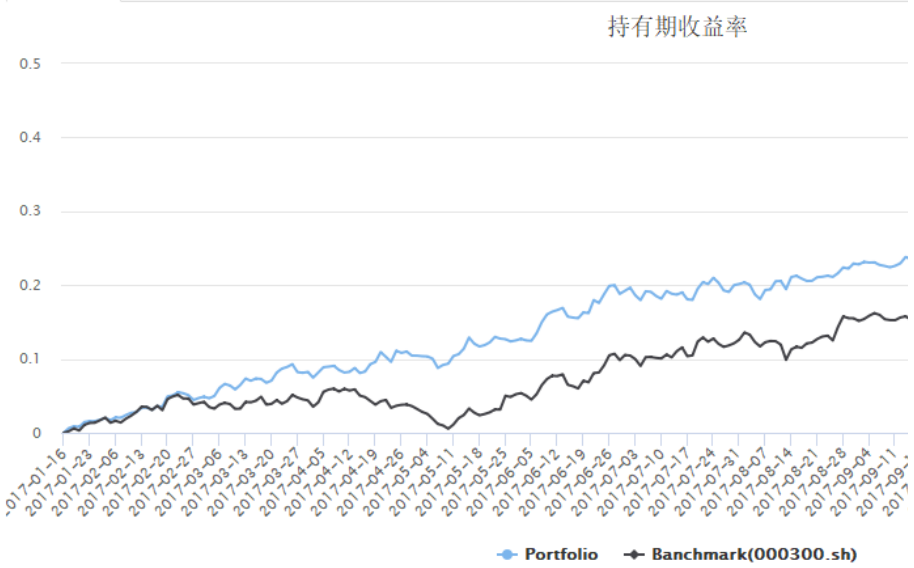
R语言量化投资常用包

- 评价和风险指标：PerformanceAnalytics
- 策略账户组合：quantstrat，blotter，ProfolioAnalytics
- 技术指标计算：TTR, quantmod
- 时间序列类型： zoo, xts



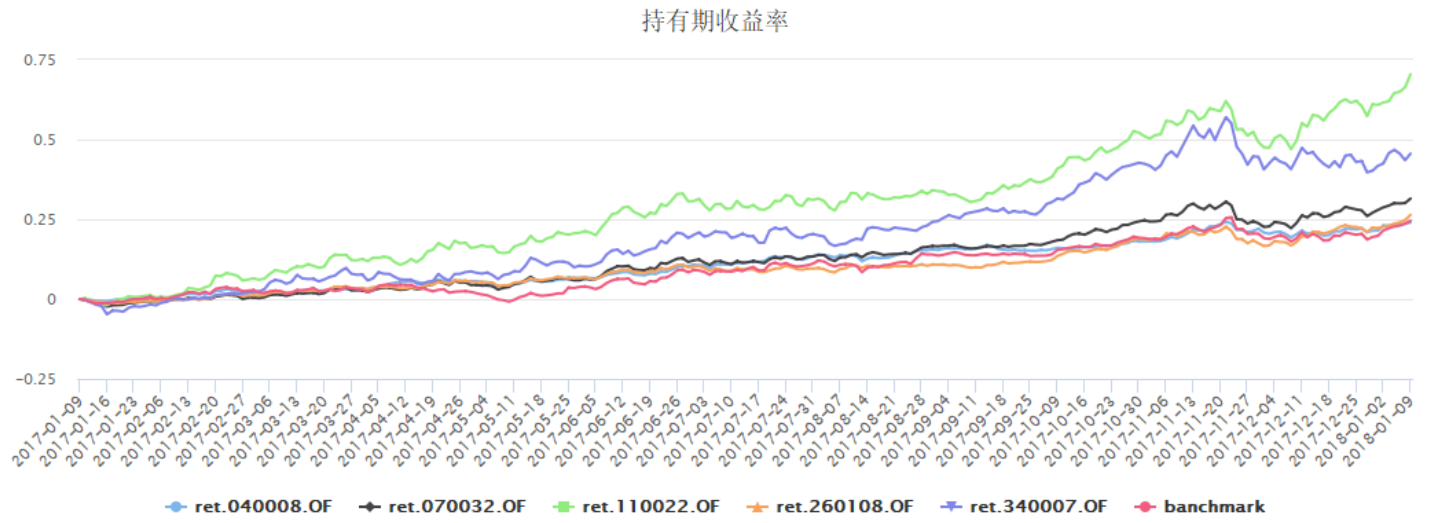
投资组合回测

组合表现 基金成分



	持有期	近1周	近1月	近3月	近6月	近1年
累计收益率(算术平均)	0.3533	0.0158	0.0414	0.0989	0.1792	0.3533
累计收益率(几何平均)	0.4237	0.0159	0.0423	0.1039	0.1963	0.4237
波动率	0.0065	0.0059	0.0069	0.0091	0.0074	0.0065
Beta	0.8114	1.6194	0.8808	0.9978	0.8776	0.8114
Alpha	0.2023	-0.0175	-0.0017	0.0265	0.0701	0.1947
无风险收益率	0.0389	0.0008	0.0033	0.0100	0.0200	0.0400
最大回撤	-0.0697	-0.0030	-0.0236	-0.0697	-0.0697	-0.0697
Sharpe	0.1999	0.4192	0.2665	0.1548	0.1756	0.1999

组合表现 基金成分



基本信息 累计收益率 波动率 Beta Alpha 最大回撤 Sharpe

	基金代码	基金简称	基金成立日	基金管理人	基金托管	管理费率%	托管费率%
1	040008.OF	华安策略优选	2007-08-02	华安基金管理有限公司	交通银行	1.5000	0.2500
2	070032.OF	嘉实优化红利	2012-06-26	嘉实基金管理有限公司	中国银行	1.5000	0.2500
3	110022.OF	易方达消费行业	2010-08-20	易方达基金管理有限公司	中国农业银行	1.5000	0.2500
4	260108.OF	景顺长城新兴成长	2006-06-28	景顺长城基金管理有限公司	中国工商银行	1.5000	0.2500
5	340007.OF	兴全社会责任	2008-04-30	兴全基金管理有限公司	中国建设银行	1.5000	0.2500

PART

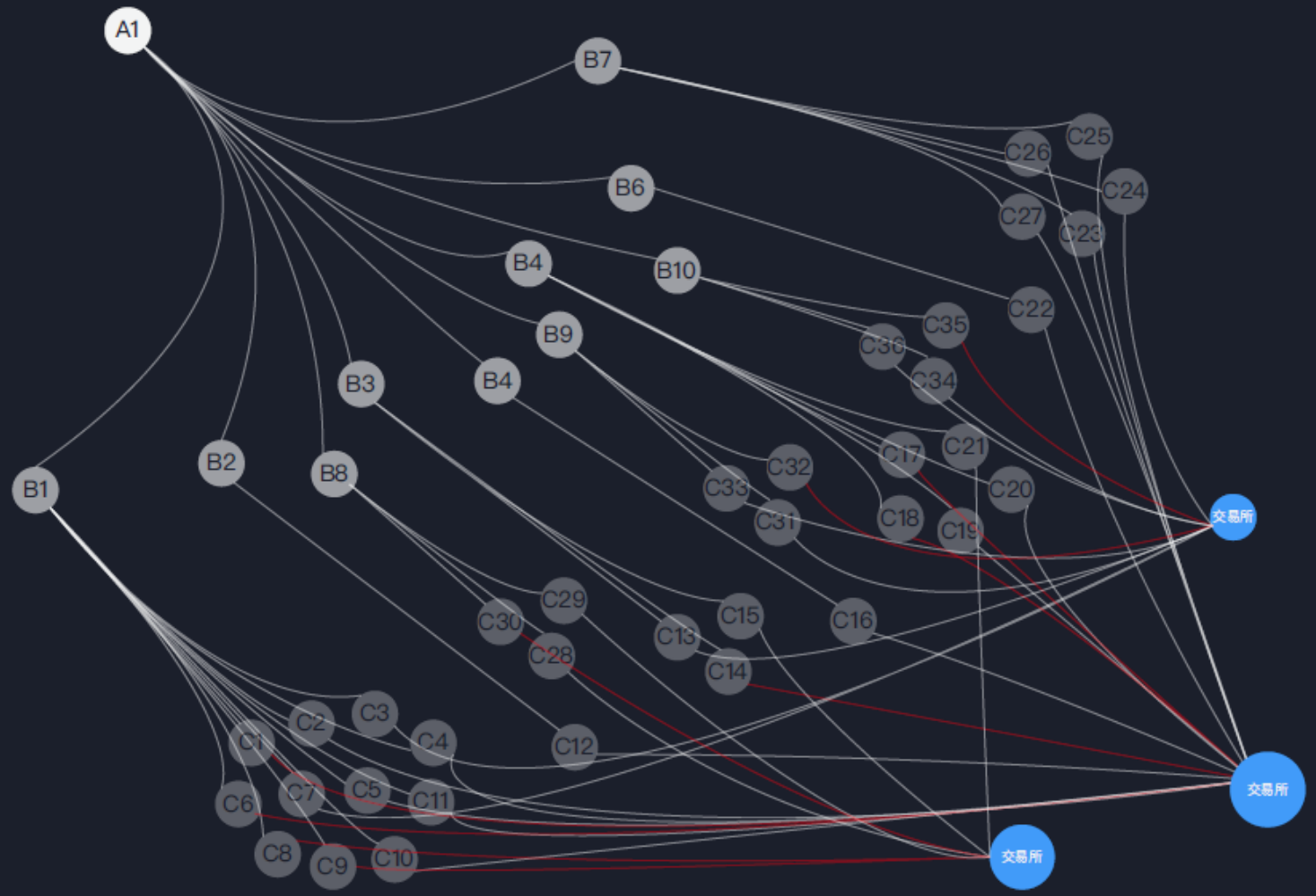
03

区块链

链上数据，让交易无所遁形

案例分析

让可疑交易无所遁形



A1 0x55f93985431fc9304077687a35a1ba103dc1e081
B1 0xdf31a499a5a8358b74564f1e2214b31bb34eb46f
B2 0xd6a09bdb29e1eafa92a30373c44b09e2e2e0651e
B3 0x5faf679d017438f481faf55e0f0cac3998ed1c88
B4 0x80c0f1ba0e243f5c5c8cc45e0d6c95ce1cbff41b
B5 0xe29526b6d7def7433823851c25f36a365567165f
B6 0x38c1617864cbab17a5e4078ffd37a587ef0d4236
B7 0xb4d30cac5124b46c2df0cf3e3e1be05f42119033
C1 0xd6a09bdb29e1eafa92a30373c44b09e2e2e0651e
C2 0xdf31a499a5a8358b74564f1e2214b31bb34eb46f
C3 0xe29526b6d7def7433823851c25f36a365567165f
C4 0x80c0f1ba0e243f5c5c8cc45e0d6c95ce1cbff41b
C5 0x843947113621901dd913377424e967e81a97ff79
C6 0xe999f794a8b5c6ec6e62076a6d9ec4c837d88bba
C7 0xb1bdd358f42ecd68ab82351f6064b26002bc60ff
C8 0x38c1617864cbab17a5e4078ffd37a587ef0d4236
C9 0xce1bf2889ee155e25b3c6d685cb05f454a4f36ba
C10 0x56e8399f5a1f6e5fba4200a1c4961b1677d03f24
C11 0x9a0c0bbd3f4773473057a32c02bd25c63196bcd6
C12 0xb4d30cac5124b46c2df0cf3e3e1be05f42119033
C13 0x5faf679d017438f481faf55e0f0cac3998ed1c88
C14 0x56e8399f5a1f6e5fba4200a1c4961b1677d03f24
C15 0xe28a30444316a74cdf3abc66c8075cbeb726fb9a
C16 0xb1bdd358f42ecd68ab82351f6064b26002bc60ff
C17 0x5faf679d017438f481faf55e0f0cac3998ed1c88
C18 0xe9a9a423bb7da5c56a06272670c3e2b20709bdb2
C19 0xf65b227378b1be547e6f2f2147692c49f1bff149
C20 0x295469fbfe3b01a6e6b206934ad044239affd2e3
C21 0x376ca9340b3033d739647c56448a8c50c2425dd2
C22 0xc9ef35f680bc03dfaab2f7f6215642fd1ece1af5
C23 0xe5984fe3855e53de863d2573417d35cffaebb475
C24 0xa82747e502b62b1b7743954aed091a1f34395cf6
C25 0x3aa6b47e17d88c72e77ee560b16dcd6068428610
C26 0x38c1617864cbab17a5e4078ffd37a587ef0d4236
C27 0xb4d30cac5124b46c2df0cf3e3e1be05f42119033

SOC

TOKEN All Sports Coin

总供给量：1,500,000,000，　2018-01-19

root钱包：0x098b46cf0b8b1ee142214d1446fa4828133e94c9　0

数据时间：2018-07-13

钱包统计：

- 25995(累积),
- 3908(无币),
- 22087(有币),
- 17943(小于10),
- 4144(大于10)

数据统计到：593872 block

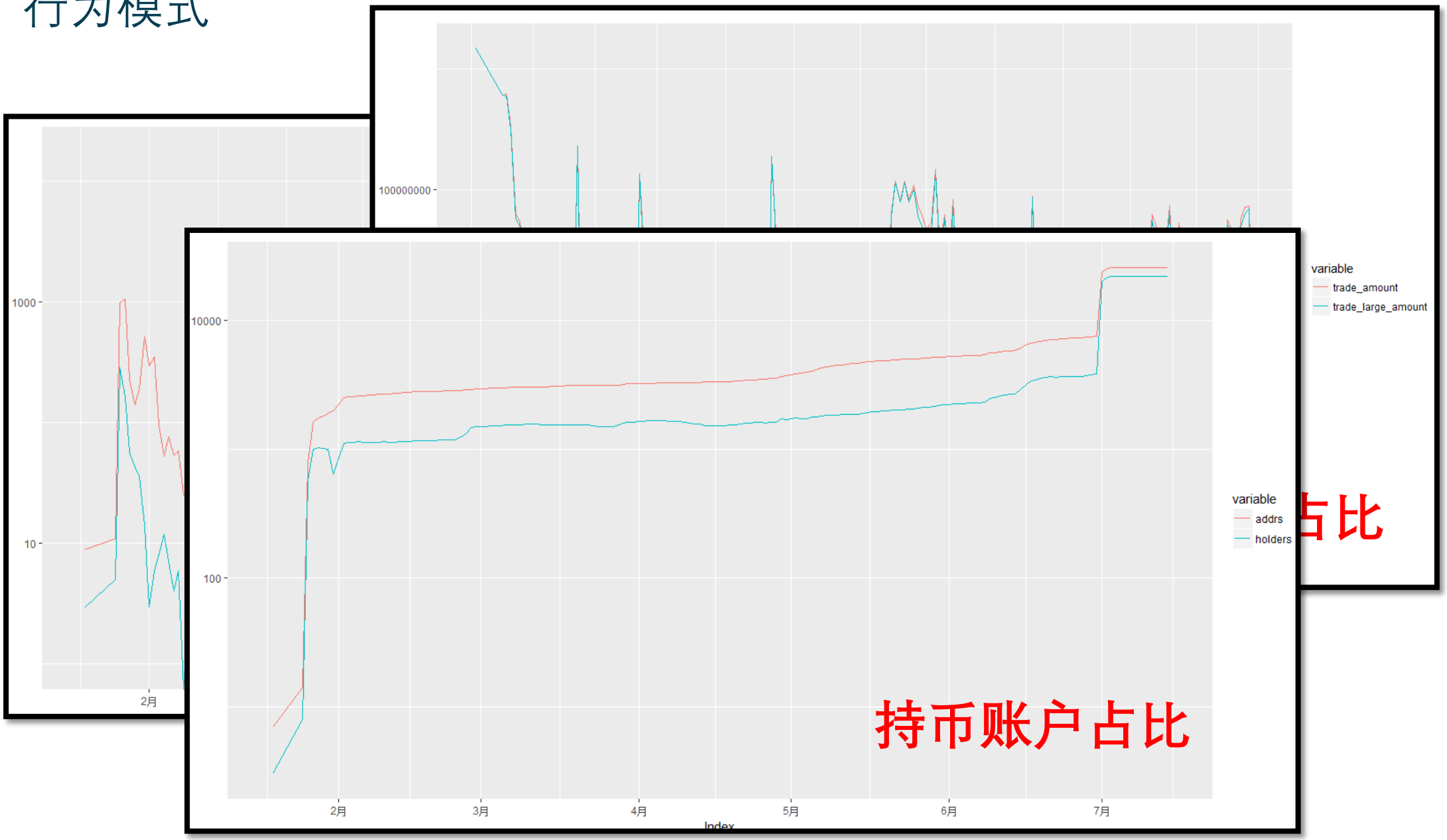
筹码分布：排除root钱包

	label	size	amount	percent
1	top_10	10	1183473250	0.78898
2	top_50	50	1375261490	0.91684
3	top_100	100	1428028981	0.95202
4	top_r1	41	1356437805	0.90429
5	top_r5	207	1459865415	0.97324
6	top_r10	414	1481859706	0.98791

币流与钱包

日期	交易数	交易量	单交易数	大单交易量	累积钱包数	持币钱包数
2018-01-19	9	1500060000	3	1500000000	7	3
2018-01-25	11	614625200	5	614595198	14	8
2018-01-26	989	617871945	286	593141186	791	569
2018-01-27	1047	339392903	164	317861237	1624	995
2018-01-28	227	64676881	56	57245320	1761	1038
2018-01-29	140	54270400	43	50697397	1813	1007
2018-01-30	192	33574037	36	29187615	1901	996
2018-01-31	514	21948757	14	16481561	1982	644
2018-02-01	297	5860269	3	3907250	2218	858
2018-02-02	350	5870640	6	4787950	2511	1113
2018-02-03	96	9012032	8	7845860	2548	1125
2018-02-04	53	19477472	12	18480920	2570	1136
2018-02-05	76	7547339	7	5477112	2604	1147
2018-02-06	54	2265603	4	1117382	2622	1134
2018-02-07	58	3120328	6	1602370	2646	1133
2018-02-08	25	362714	0	0	2658	1135
2018-02-09	26	2084572	2	2000000	2670	1140
2018-02-10	28	2942590	4	2310178	2689	1151
2018-02-11	76	232065129	9	230560630	2703	1141
2018-02-12	94	3298059	6	1468580	2722	1139
2018-02-13	60	14047869	5	12853165	2746	1153
2018-02-14	38	7258376	11	6353263	2762	1157
2018-02-15	44	12507077	9	11714880	2781	1152

行为模式


100000000
1000
10
2月
10000
100
2月
3月
4月
5月
6月
7月
Index
variable
trade_amount
trade_large_amount
variable
addrs
holders
占比
持币账户占比

BNB

TOKEN BNB

总供给量：194,972,068　　2017-07-08

root钱包：0x00c5e04176d95a286fcce0e68c683ca0bfec8454　　92833406　　0.47614

数据时间：2018-07-12

钱包统计：

- 324011(累积),
- 22350(无币),
- 301661(有币),
- 289641(小于10),
- 12020(大于10)

筹码分布：排除root钱包

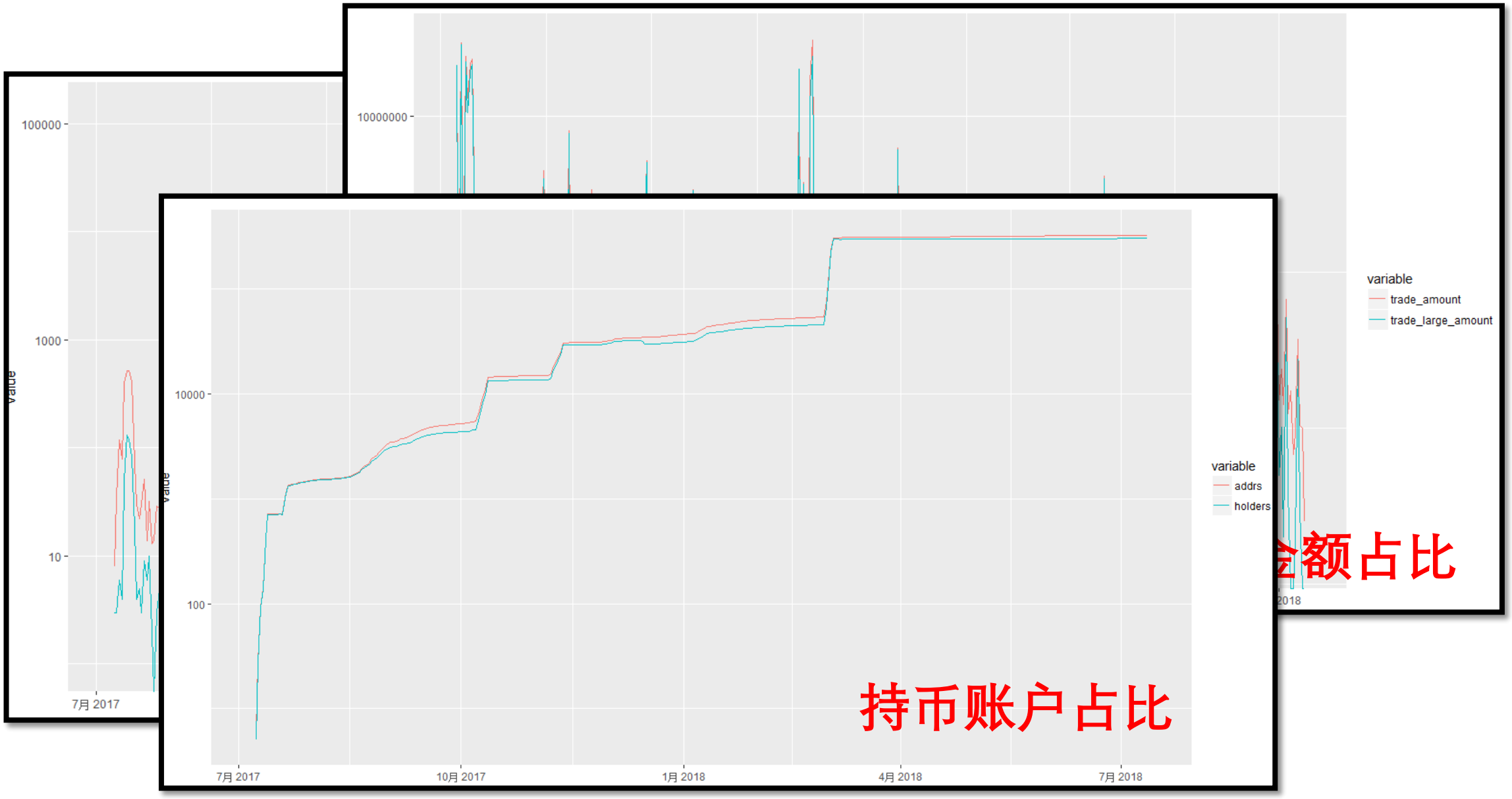
	label	size	amount	percent
1	top_10	10	75634056	0.38792
2	top_50	50	84631708	0.43407
3	top_100	100	88975329	0.45635
4	top_r1	3016	101699644	0.52161
5	top_r5	15083	102093510	0.52363
6	top_r10	30166	102114121	0.52374

数据统计到：5942634 block

币流与钱包

日期	交易数	交易量	大单交易数	大单交易量	积钱包数	持币钱包数
2017-07-08	8	21069474	3	21069430	6	5
2017-07-09	35	396755	3	231169	28	27
2017-07-10	119	29449102	6	28707999	94	91
2017-07-11	79	718316	4	247666	140	137
2017-07-12	401	24364524	49	22388209	414	411
2017-07-13	522	13055930	131	10599657	715	702
2017-07-14	524	21666417	120	19268790	715	702
2017-07-15	416	23286787	84	21599969	715	702
2017-07-16	72	1762837	18	1473598	715	703
2017-07-17	30	313385	4	162587	715	704
2017-07-18	22	586015	5	518380	715	710
2017-07-19	30	341849	3	178835	715	702
2017-07-20	51	910030	9	717671	1035	1020
2017-07-21	14	510359	6	444552	1312	1295
2017-07-22	32	2916725	10	2796661	1353	1336
2017-07-23	13	291138	2	239717	1370	1352
2017-07-24	14	55986	0	0	1385	1367
2017-07-25	29	295760	3	208788	1402	1384
2017-07-26	28	358688	5	250879	1425	1407
2017-07-27	11	63564	1	20800	1434	1417
2017-07-28	15	106240	2	50060	1455	1438

行为模式


100000
1000
10
value
7月 2017
10000000
variable
trade_amount
trade_large_amount
2018
额占比
10000
100
addrs
holders
7月 2017
10月 2017
1月 2018
4月 2018
7月 2018
持币账户占比

理论，场景，价值

市场，是检验产品的唯一标准！

本次分享仅在于**数据研究**，不构成任何金融资产的推荐和建议。

张丹，青萌数海CTO，微软MVP，资深R语言技术专家。10年以上互联网应用架构经验，在R、Java、NodeJS、大数据、统计、数据挖掘算法等方面有深厚的积累。

金融大数据专家，精通量化投资交易策略，熟悉中国金融二级市场、交易规则和投研体系。熟悉数据学科方法论，在外汇领域和区块链领域做落地的尝试。多次在互联网技术、数据科学相关技术大会中担任演讲嘉宾。

著有《R的极客理想：量化投资篇》、《R的极客理想：工具篇》、《R的极客理想：高级开发篇》，英文版图书被CRC出版集团引进，在美国发行。个人博客：http://fens.me（Alexa全球排名70k。）

- 联系方式:
- Blog：http://fens.me
- Email: bsspirit@gmail.com

如果你和我一样，欢迎加入我的团队!