### 数据集1：NFL Play-by-Play 2009-2017

#### 数据摘要：

**1.标称属性：**Date、Drive、qrt、down、time、 SideofField、posteam、DefensiveTeam，

Desc、PlayAttempted、ExPointResult、PlayType、Passer、Passer\_ID、PassLocation、Interceptor、Rusher、Rusher\_ID、RunLocation、RunGap、Receiver、Receiver\_ID、ReturnResult、Returner、BlockingPlayer、Tackler1、Tackler2、FieldGoalResult、RecFumbTeam、RecFumbPlayer、PenalizedTeam、PenaltyType、PenalizedPlayer、HomeTeam、AwayTeam、Timeout\_Team、Season

由于有些标称属性取值的重复次数很低，将标称属性取值的频数展示出来很麻烦，所以这里只展示取值重复次数高的属性

标称属性取值的频数：

（1）Date：

2013-09-09:378 2013-09-08:2274 2015-11-12:169 2013-09-05:216 2015-11-16:175

2013-12-29:2729 2014-09-18:172 2013-12-23:170 2014-09-14:2465 2014-09-11:166

2015-09-10:174 2015-09-13:2291 2015-09-14:357 2015-09-17:182 2010-11-11:179

2017-11-16:187 2017-11-13:171 2017-11-12:2132 2017-11-19:2115 2015-11-19:176

2015-12-14:180 2010-11-18:158 2015-12-17:181 2015-12-10:175 2015-12-13:2456

2015-12-19:171 2012-09-13:169 2010-09-13:336 2009-12-06:2485 2009-09-21:163

2009-12-03:165 2015-11-15:2187 2010-10-24:2306 2010-10-25:188 2011-09-08:189

2016-09-08:173 2016-12-04:2271 2016-12-05:161 2016-12-01:175 2012-12-10:185

2016-12-08:174 2009-11-23:173 2014-09-15:187 2009-11-26:515 2013-12-22:2587

2009-11-29:2090 2016-11-21:173 2016-11-20:2124 2016-11-27:2171 2016-11-24:474

2014-10-26:2341 2014-10-27:173 2016-11-28:173 2014-10-23:173 2014-10-20:175

2014-11-16:2086 2014-11-17:150 2014-11-10:192 2014-11-13:169 2016-10-24:178

2014-12-18:168 2010-11-14:2215 2010-11-15:175 2014-12-11:180 2014-12-15:177

2014-12-14:2457 2013-12-12:165 2017-12-07:166 2013-12-16:169 2013-12-15:2510

2011-10-03:189 2011-10-02:2636 2011-10-09:2122 2017-11-26:2116 2017-11-27:184

2017-11-23:511 2017-11-20:184 2013-10-28:159 2013-10-24:172 2013-10-27:1924

2013-10-21:184 2013-10-20:2318 2010-12-16:160 2010-12-13:382 2010-12-12:2193

2010-12-19:2472 2016-10-06:196 2010-09-09:159 2012-12-02:2529 2012-12-03:158

2012-12-06:173 2012-12-09:2470 2017-12-04:174 2013-11-11:165 2013-11-10:2066

2013-11-17:2349 2017-12-03:2403 2013-11-14:165 2013-11-18:157 2017-10-15:2169

2017-10-16:171 2017-10-12:189 2017-10-19:188 2011-12-11:2503 2011-12-12:182

2009-11-19:181 2011-12-15:159 2012-11-29:180 2011-12-17:152 2009-11-12:168

2011-12-19:171 2012-11-25:2085 2009-11-16:166 2012-11-22:600 2009-11-15:2333

2016-10-31:164 2016-10-30:2093 2015-11-29:2187 2015-11-26:532 2013-09-30:172

2015-11-22:2129 2015-11-23:185 2014-09-21:2465 2014-09-22:181 2014-09-25:178

2014-09-28:1981 2014-09-29:154 2016-12-19:186 2011-11-28:189 2011-11-20:2109

2011-11-21:174 2016-11-10:166 2011-11-24:501 2016-11-17:185 2016-11-14:175

2011-11-27:2145 2016-12-11:2441 2015-12-03:183 2015-12-07:167 2015-12-06:2479

2009-10-04:2298 2009-10-05:167 2009-09-13:2211 2009-12-19:188 2009-09-10:178

2009-09-14:355 2009-12-13:2429 2009-12-10:163 2009-12-17:173 2009-12-14:169

2010-01-03:2716 2015-11-09:180 2016-10-09:2081 2010-10-31:2132 2016-10-02:2377

2016-10-03:184 2012-11-15:172 2012-11-12:174 2012-11-11:2122 2012-11-18:2227

2012-11-19:156 2016-09-12:348 2016-09-11:2316 2016-12-18:2259 2016-09-15:175

2016-12-12:192 2016-09-19:171 2016-09-18:2589 2016-12-17:164 2016-12-15:159

2012-09-09:2294 2012-09-05:158 2013-11-28:498 2013-11-21:165 2013-11-24:2196

2013-11-25:156 2012-10-08:165 2012-10-01:166 2012-10-07:2127 2012-10-04:172

2017-09-25:176 2017-09-24:2508 2017-09-21:196 2014-10-30:171 2017-09-28:166

2010-11-07:2109 2014-12-28:2771 2010-11-01:185 2014-12-20:401 2014-12-21:2128

2014-12-22:182 2010-11-08:166 2013-12-01:2133 2013-12-02:169 2013-12-05:200

2013-12-09:165 2013-12-08:2390 2011-10-30:2150 2011-10-31:198 2009-12-07:185

2017-11-30:167 2013-10-10:160 2013-10-13:2324 2013-10-14:165 2013-10-17:186

2010-12-20:171 2010-12-23:165 2010-12-25:180 2010-12-26:2083 2010-12-27:172

2014-11-09:1943 2014-11-03:208 2014-11-02:1968 2014-11-06:187 2010-09-19:2565

2009-09-28:164 2009-09-20:2674 2010-09-12:2294 2009-09-27:2575 2012-12-17:169

2012-12-16:2468 2009-12-20:2296 2009-12-21:159 2012-12-13:185 2009-12-27:2455

2009-12-25:162 2017-12-31:2801 2009-12-28:197 2017-09-18:154 2011-09-25:2663

2011-09-26:174 2017-10-09:182 2016-09-29:158 2015-10-26:169 2016-09-22:173

2017-10-02:171 2017-10-01:2523 2016-09-26:184 2016-11-03:193 2017-10-05:193

2016-09-25:2506 2016-12-22:189 2016-12-26:172 2016-12-24:2192 2016-12-25:364

2009-11-09:153 2009-11-08:2127 2011-12-05:165 2011-12-04:2557 2017-09-10:2051

2011-12-01:152 2009-11-01:2070 2009-11-02:191 2011-12-08:162 2013-09-23:178

2013-09-22:2478 2015-11-30:175 2013-09-26:187 2013-09-29:2374 2016-10-20:177

2010-12-05:2382 2015-10-29:180 2016-01-03:2872 2015-10-25:2183 2015-10-22:157

2016-11-07:183 2016-11-06:1988 2014-11-30:2147 2009-10-19:168 2009-10-18:2282

2009-10-11:2241 2009-10-12:174 2017-09-17:2472 2012-01-01:2825 2017-09-14:178

2010-10-04:167 2016-10-17:173 2016-10-16:2344 2016-10-10:174 2016-10-13:183

2010-10-03:2215 2012-11-05:167 2012-11-04:2130 2012-11-01:146 2012-11-08:175

2011-12-18:2260 2012-11-26:161 2012-09-17:186 2012-09-16:2477 2012-12-22:177

2012-12-23:2612 2012-09-10:347 2011-11-06:2277 2011-11-07:176 2015-10-15:193

2015-10-12:177 2015-10-11:2183 2015-10-18:2254 2015-10-19:191 2012-10-18:150

2012-10-11:178

2010-10-17:2340

2012-10-15:176

2012-10-14:2207

2017-10-30:186

2009-11-22:2437

2014-10-02:175

2014-10-05:2348

2014-10-06:176

2014-10-09:184

2012-09-30:2319

2017-09-11:336

2015-11-08:2064

2013-09-19:192

2015-11-02:215

2015-11-01:2138

2013-09-16:183

2015-11-05:156

2014-09-08:340

2014-09-07:2392

2014-09-04:171

2015-09-28:196

2015-09-24:190

2015-09-27:2488

2015-09-21:167

2015-09-20:2587

2013-09-12:186

2017-11-02:194

2017-11-05:1929

2017-11-06:161

2017-11-09:187

2013-10-07:158

2013-10-06:2128

2013-10-03:200

2015-12-21:179

2015-12-20:2343

2015-12-24:208

2015-12-27:2222

2015-12-26:211

2015-12-28:183

2010-09-26:2644

2010-09-27:159

2010-09-20:172

2016-10-23:2357

2017-12-25:329

2017-12-24:2153

2017-12-23:344

2011-10-23:2080

2011-10-24:173

2011-09-18:2591

2011-09-19:188

2011-11-14:173

2011-09-11:2246

2011-09-12:395

2012-09-27:204

2012-09-24:172

2017-01-01:2819

2012-09-23:2607

2012-09-20:160

2011-01-02:2772

2009-11-30:161

2013-09-15:2544

2016-11-13:2153

2017-09-07:203

2014-11-20:181

2014-11-23:2311

2014-11-24:175

2014-11-27:519

2014-12-08:180

2010-11-29:162

2010-11-28:2083

2010-11-25:542

2014-12-01:162

2010-11-21:2433

2014-12-07:2484

2014-12-04:174

2010-11-22:174

2009-10-25:2061

2009-10-26:178

2010-12-28:177

2011-10-10:172

2011-10-16:2116

2011-10-17:159

2017-10-08:2144

2013-10-31:198

2010-12-06:164

2010-10-11:193

2010-10-10:2304

2010-12-02:179

2016-10-27:186

2010-10-18:166

2010-12-09:178

2012-12-30:2737

2017-12-16:344

2017-12-17:2133

2017-12-14:175

2017-12-10:2496

2017-12-11:186

2013-11-04:172

2013-11-07:173

2017-12-18:169

2013-11-03:1961

2015-10-05:158

2015-10-04:2402

2011-11-17:177

2015-10-01:192

2011-11-10:175

2011-11-13:2426

2015-10-08:180

2012-10-28:2116

2012-10-29:171

2012-10-25:191

2012-10-21:1957

2012-10-22:188

2017-10-23:176

2017-10-22:2258

2017-10-26:177

2017-10-29:1963

2014-10-16:188

2014-10-13:189

2014-10-12:2397

2011-12-25:164

2011-12-24:2288

2011-12-26:183

2014-10-19:2260

2011-12-22:184

1. Drive

24:7959 25:5842 26:4200 27:2815 20:15622 21:13998 22:12071 23:9933

28:1658 29:1077 1:19261 3:17425 2:17084 5:17822 4:17875 7:18088 6:17925 9:17033

8:17691 11:16898 10:17193 13:16745 12:16724 15:17188 14:16881

17:17459 16:17525 19:16928 18:17339 31:366 30:675 35:9 34:45

33:97 32:237

1. qrt

1:89176 3:90682 2:112317 5:2872 4:112641

（4）Down

1:138878 NA:61154 3:67398 2:104089 4:36169

（5）SideofField

MIN:12277 MIA:12924 CAR:12222 ATL:12088 DET:12799 MID:4326 NYJ:12396 DEN:12606 BAL:12442 NYG:12855 OAK:13270 TEN:12968 NO:12566 DAL:12306

NA:528 NE:12001 BUF:13094 SEA:12077 CHI:12583 TB:12788 JAX:2754

STL:9879 CLE:13212 CIN:12384 HOU:12628 GB:12250 WAS:12674 JAC:10185

KC:12486 PHI:12673 PIT:12023 LA:2763 50:312 LAC:1351 IND:12760

ARI:12834 SF:12452 SD:10952

（6）posteam

:24992 MIN:11774 MIA:11690 CAR:11870 ATL:12004 DET:12271 CIN:11931

NYJ:12076 DEN:12398 BAL:12241 NYG:12021 OAK:12002 TEN:11511

NO:12408 DAL:11784 NE:12520 BUF:11767 SEA:11764 CHI:11566

TB:11696 JAX:2692 STL:9113 CLE:11722 HOU:12315 GB:12052

WAS:11905 JAC:9253 KC:11808 PHI:12397 PIT:11859 LA:2599

LAC:1312 IND:12156 ARI:11967 SF:11591 SD:10661

1. DefensiveTeam

MIN:11911 MIA:12133 CAR:11646 ATL:11623 DET:11965 CIN:11945

NYJ:11801 DEN:12222 BAL:12025 NYG:12089 OAK:12158 TEN:12293

NO:11890 DAL:11769 NA:24992 NE:12073 BUF:12237 SEA:11832

CHI:11772 TB:11890 JAX:2594 STL:9290 CLE:12228 HOU:11812

GB:11759 WAS:11888 JAC:9401 KC:12020 PHI:12321 PIT:11499

LA:2702 LAC:1285 IND:12014 ARI:12236 SF:12109 SD:10264

（8）PlayAttempted

1:407688

（9）ExPointResult

NA:397578 Made:9850 Missed:172 Aborted:7 Blocked:81

1. PlayType

Quarter End:4914 Half End:40 QB Kneel:3530 Run:120831 Spike:640

Extra Point:10063 Sack:10649 Field Goal:8928 Punt:22003 Kickoff:23403

Timeout:16206 Pass:159353 Two Minute Warning:3741 End of Game:1973

No Play:21414

1. PassLocation

middle:38992 NA:240520 right:67675 left:60501

1. RunLocation

NA:288178 right:43532 left:43157 middle:32821

（13）RunGap

tackle:29089 NA:320260 guard:27074 end:31265

1. ReturnResult

Touchdown:1001 NA:389450 Fair Catch:5552 Touchback:11685

1. FieldGoalResult

NA:398629 Blocked:205 Good:7582 No Good:1272

1. RecFumbTeam

MIN:135 MIA:123 CAR:142 ATL:138 DET:123 CIN:111 NYJ:145 DEN:127 BAL:140 NYG:159

TEN:114 NO:131 DAL:146 NA:403315 NE:132 SEA:143 CLE:125 BUF:133 PIT:127

STL:107 OAK:143 HOU:134 GB:130 CHI:165 WAS:139 JAC:147 KC:134 PHI:146 TB:161

LA:35 LAC:14 IND:128 ARI:154 SF:136 SD:106

（16）PenalizedTeam

MIN:858 MIA:863 CAR:852 ATL:777 DET:1007 CIN:915 NYJ:897 DEN:977 BAL:975

NYG:823 OAK:1159 TEN:932 NO:895 DAL:978 NA:378189 NE:822 BUF:966

SEA:1080 CHI:920 TB:1010 STL:807 CLE:906 HOU:916 GB:851 WAS:904 JAC:846

KC:892 PHI:990 PIT:863 LA:231 LAC:107 IND:797 ARI:917 SF:983 SD:783

（17）HomeTeam

MIN:12619 MIA:12758 CAR:12551 ATL:12431 DET:12813 CIN:12675

NYJ:12706 DEN:13041 BAL:12902 NYG:12901 OAK:12895 TEN:12689

NO:12984 DAL:12597 NE:13165 BUF:12996 SEA:12559 CHI:12516

TB:12507 JAX:2723 STL:9842 CLE:12858 HOU:12731 GB:12743

WAS:12555 JAC:10035 KC:12620 PHI:13123 PIT:12624 LA:2798

LAC:1363 IND:12897 ARI:12897 SF:12567 SD:11007

1. AwayTeam

MIN:12553 MIA:12685 CAR:12496 ATL:12743 DET:12975 CIN:12763

NYJ:12668 DEN:13127 BAL:12927 NYG:12704 OAK:12815 TEN:12603

NO:13003 DAL:12553 NE:12959 SEA:12530 CLE:12700 BUF:12705

PIT:12278 JAX:2930 STL:9663 CHI:12562 HOU:12939 GB:12666

WAS:12809 JAC:9808 KC:12824 PHI:13129 TB:12646 LA:2836

LAC:1415 IND:12780 ARI:12857 SF:12735 SD:11302

1. Timeout\_Team

MIN:446 MIA:545 CAR:534 ATL:484 DET:527 CIN:537 NYJ:505 DEN:544 BAL:557

NYG:498 OAK:558 TEN:486 NO:628 DAL:541 NE:475 BUF:501 SEA:478 CHI:519

TB:541 None:390885 JAX:103 STL:360 CLE:573 HOU:546 GB:530 WAS:555 JAC:416

KC:580 PHI:525 PIT:523 LA:99 LAC:59 IND:462 ARI:507 SF:572 SD:489

1. Season

2016:45736 2015:46129 2014:45502 2017:45241 2009:44596 2011:45070

2010:44751 2013:45248 2012:45415

**2.数值属性：**TimeUnder、TimeSecs、PlayTimeDiff、yrdln、yrdline100、ydstogo、ydsnet、AirYards、YardsAfterCatch、FieldGoalDistance、Penalty.Yards、PosTeamScore、DefTeamScore、ScoreDiff、AbsScoreDiff、posteam\_timeouts\_pre、HomeTimeouts\_Remaining\_Ere、AwayTimeouts\_Remaining\_Pre、HomeTimeouts\_Remaining\_Post、AwayTimeouts\_Remaining\_Post、No\_Score\_Prob、Opp\_Field\_Goal\_Prob、Opp\_Safety\_Prob、Opp\_Touchdown\_Prob、Field\_Goal\_Prob、Safety\_Prob、Touchdown\_Prob、ExPoint\_Prob、TwoPoint\_Prob、ExpPts、EPA、airEPA、yacEPA、Home\_WP\_pre、Away\_WP\_pre、Home\_WP\_post、Away\_WP\_post、Win\_Prob、WPA、airWPA、yacWPA、

TimeUnder:

最大值:15.00 最小值:0.00 均值:0.00 中位数:7.00 四分位数\_1:3.00 四分位数\_3:11.00 缺失值:0

TimeSecs:

最大值:3600.00 最小值:-900.00 均值:-900.00 中位数:1800.00 四分位数\_1:778.00 四分位数\_3:2585.00 缺失值:224

PlayTimeDiff:

最大值:943.00 最小值:0.00 均值:0.00 中位数:17.00 四分位数\_1:5.00 四分位数\_3:37.00 缺失值:444

yrdln:

最大值:50.00 最小值:1.00 均值:1.00 中位数:30.00 四分位数\_1:20.00 四分位数\_3:39.00 缺失值:840

yrdline100:

最大值:99.00 最小值:1.00 均值:1.00 中位数:49.00 四分位数\_1:30.00 四分位数\_3:70.00 缺失值:840

ydstogo:

最大值:50.00 最小值:0.00 均值:0.00 中位数:9.00 四分位数\_1:3.00 四分位数\_3:10.00 缺失值:0

ydsnet:

最大值:99.00 最小值:-87.00 均值:-87.00 中位数:19.00 四分位数\_1:5.00 四分位数\_3:43.00 缺失值:0

AirYards:

最大值:84.00 最小值:-70.00 均值:-70.00 中位数:0.00 四分位数\_1:0.00 四分位数\_3:4.00 缺失值:0

YardsAfterCatch:

最大值:90.00 最小值:-81.00 均值:-81.00 中位数:0.00 四分位数\_1:0.00 四分位数\_3:0.00 缺失值:0

FieldGoalDistance:

最大值:71.00 最小值:18.00 均值:18.00 中位数:38.00 四分位数\_1:29.00 四分位数\_3:46.00 缺失值:398740

Penalty.Yards:

最大值:66.00 最小值:0.00 均值:0.00 中位数:0.00 四分位数\_1:0.00 四分位数\_3:0.00 缺失值:0

PosTeamScore:

最大值:61.00 最小值:0.00 均值:0.00 中位数:7.00 四分位数\_1:2.00 四分位数\_3:16.00 缺失值:26904

DefTeamScore:

最大值:61.00 最小值:0.00 均值:0.00 中位数:10.00 四分位数\_1:3.00 四分位数\_3:17.00 缺失值:26904

ScoreDiff:

最大值:59.00 最小值:-59.00 均值:-59.00 中位数:0.00 四分位数\_1:-7.00 四分位数\_3:4.00 缺失值:24988

AbsScoreDiff:

最大值:59.00 最小值:0.00 均值:0.00 中位数:7.00 四分位数\_1:3.00 四分位数\_3:11.00 缺失值:26904

posteam\_timeouts\_pre:

最大值:3.00 最小值:0.00 均值:0.00 中位数:3.00 四分位数\_1:2.00 四分位数\_3:3.00 缺失值:0

HomeTimeouts\_Remaining\_Pre:

最大值:3.00 最小值:-3.00 均值:-3.00 中位数:3.00 四分位数\_1:2.00 四分位数\_3:3.00 缺失值:0

AwayTimeouts\_Remaining\_Pre:

最大值:3.00 最小值:-1.00 均值:-1.00 中位数:3.00 四分位数\_1:2.00 四分位数\_3:3.00 缺失值:0

HomeTimeouts\_Remaining\_Post:

最大值:3.00 最小值:-3.00 均值:-3.00 中位数:3.00 四分位数\_1:2.00 四分位数\_3:3.00 缺失值:0

AwayTimeouts\_Remaining\_Post:

最大值:3.00 最小值:-1.00 均值:-1.00 中位数:3.00 四分位数\_1:2.00 四分位数\_3:3.00 缺失值:0

No\_Score\_Prob:

最大值:1.00 最小值:0.00 均值:0.00 中位数:0.02 四分位数\_1:0.00 四分位数\_3:0.17 缺失值:176

Opp\_Field\_Goal\_Prob:

最大值:0.36 最小值:0.00 均值:0.00 中位数:0.08 四分位数\_1:0.03 四分位数\_3:0.15 缺失值:176

Opp\_Safety\_Prob:

最大值:0.03 最小值:0.00 均值:0.00 中位数:0.00 四分位数\_1:0.00 四分位数\_3:0.00 缺失值:176

Opp\_Touchdown\_Prob:

最大值:0.50 最小值:0.00 均值:0.00 中位数:0.12 四分位数\_1:0.04 四分位数\_3:0.23 缺失值:176

Field\_Goal\_Prob:

最大值:0.99 最小值:0.00 均值:0.00 中位数:0.23 四分位数\_1:0.15 四分位数\_3:0.33 缺失值:176

Safety\_Prob:

最大值:0.02 最小值:0.00 均值:0.00 中位数:0.00 四分位数\_1:0.00 四分位数\_3:0.00 缺失值:176

Touchdown\_Prob:

最大值:0.91 最小值:0.00 均值:0.00 中位数:0.31 四分位数\_1:0.19 四分位数\_3:0.41 缺失值:176

ExPoint\_Prob:

最大值:0.99 最小值:0.00 均值:0.00 中位数:0.00 四分位数\_1:0.00 四分位数\_3:0.00 缺失值:0

TwoPoint\_Prob:

最大值:0.47 最小值:0.00 均值:0.00 中位数:0.00 四分位数\_1:0.00 四分位数\_3:0.00 缺失值:0

ExpPts:

最大值:6.50 最小值:-3.84 均值:-3.84 中位数:1.26 四分位数\_1:0.32 四分位数\_3:2.88 缺失值:176

EPA:

最大值:9.51 最小值:-13.49 均值:-13.49 中位数:0.00 四分位数\_1:-0.60 四分位数\_3:0.56 缺失值:369

airEPA:

最大值:7.35 最小值:-12.85 均值:-12.85 中位数:0.30 四分位数\_1:-0.50 四分位数\_3:1.39 缺失值:248394

yacEPA:

最大值:9.56 最小值:-14.00 均值:-14.00 中位数:0.00 四分位数\_1:-0.96 四分位数\_3:0.49 缺失值:248498

Home\_WP\_pre:

最大值:1.00 最小值:0.00 均值:0.00 中位数:0.53 四分位数\_1:0.33 四分位数\_3:0.77 缺失值:24954

Away\_WP\_pre:

最大值:1.00 最小值:0.00 均值:0.00 中位数:0.47 四分位数\_1:0.23 四分位数\_3:0.68 缺失值:24954

Home\_WP\_post:

最大值:1.00 最小值:0.00 均值:0.00 中位数:0.53 四分位数\_1:0.32 四分位数\_3:0.77 缺失值:26587

Away\_WP\_post:

最大值:1.00 最小值:0.00 均值:0.00 中位数:0.47 四分位数\_1:0.23 四分位数\_3:0.68 缺失值:26587

Win\_Prob:

最大值:1.00 最小值:0.00 均值:0.00 中位数:0.50 四分位数\_1:0.28 四分位数\_3:0.73 缺失值:25009

WPA:

最大值:0.99 最小值:-1.00 均值:-1.00 中位数:0.00 四分位数\_1:-0.01 四分位数\_3:0.01 缺失值:5541

airWPA:

最大值:0.99 最小值:-1.00 均值:-1.00 中位数:0.00 四分位数\_1:-0.01 四分位数\_3:0.04 缺失值:248501

yacWPA:

最大值:1.00 最小值:-0.99 均值:-0.99 中位数:0.00 四分位数\_1:-0.02 四分位数\_3:0.01 缺失值:248762

#### 数据可视化：

数值属性之间的相关关系为：

1. ScoreDiff=PosTeamScore–DefTeamScore

2、AbsScoreDiff= ScoreDiff的绝对值

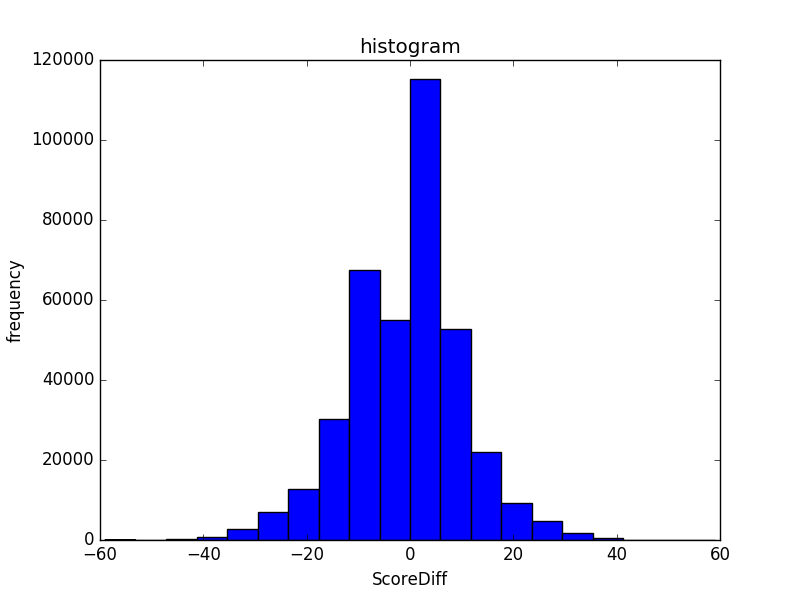
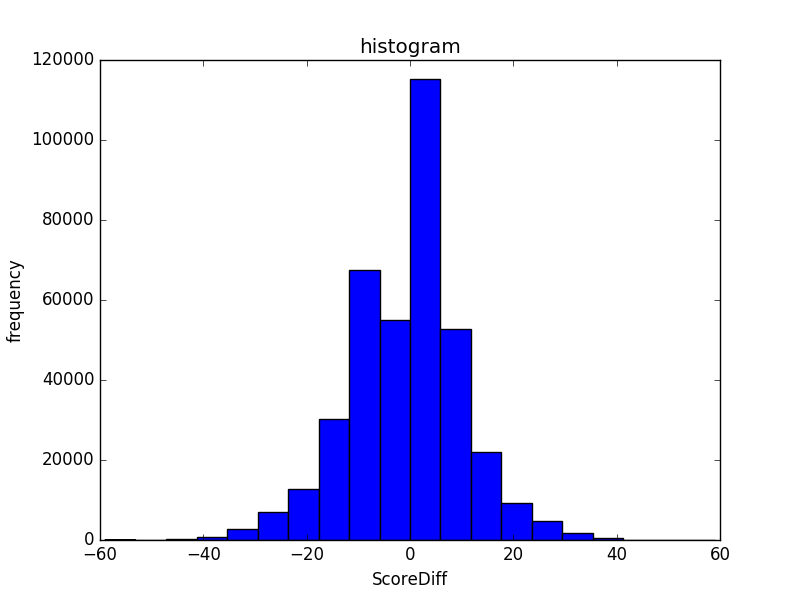
3、Home\_WP\_pre+ Away\_WP\_pre=1

4、Home\_WP\_post+ Away\_WP\_post=1

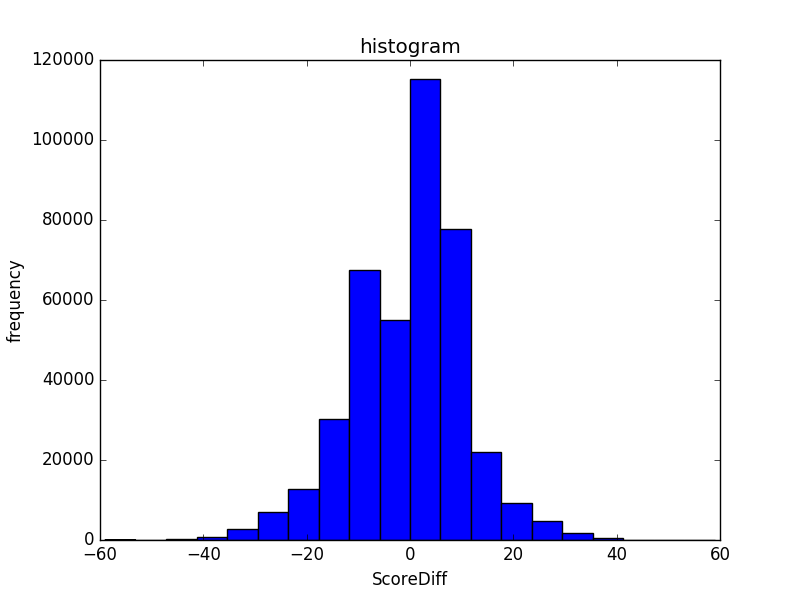
5、WPA= airWPA+yacWPA

可以从下面的图片中看到，尽管使用了不同的填充方法填充缺失值，但因为数据数量很大，而缺失值只占一小部分，所以使用不同的填充方法画出来的结果是大体相似的

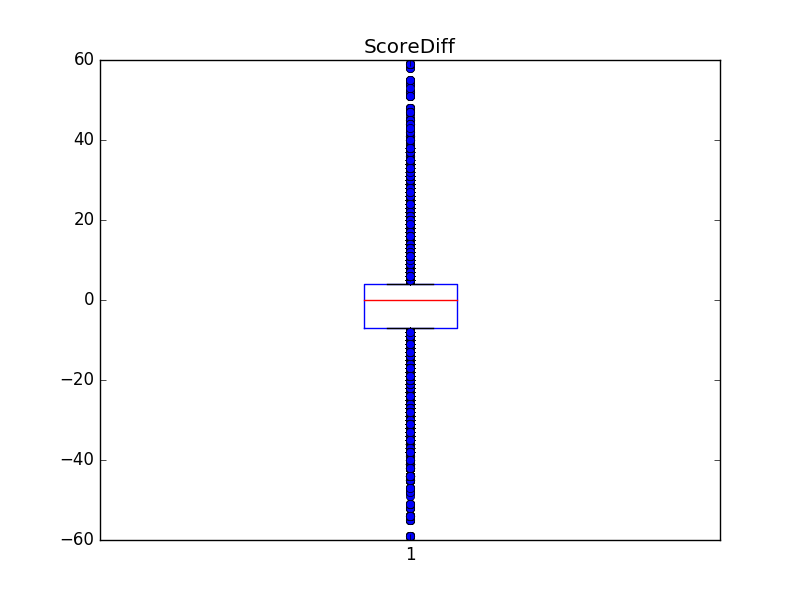
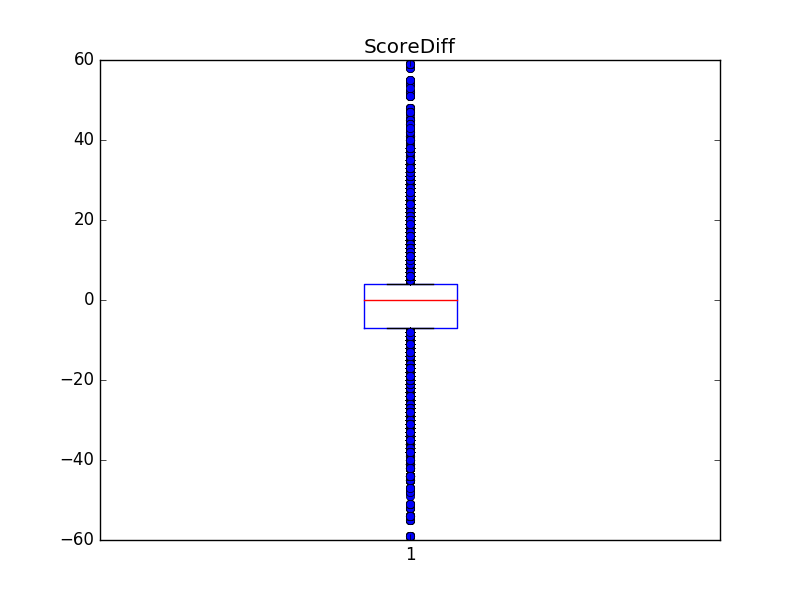
（1）ScoreDiff：



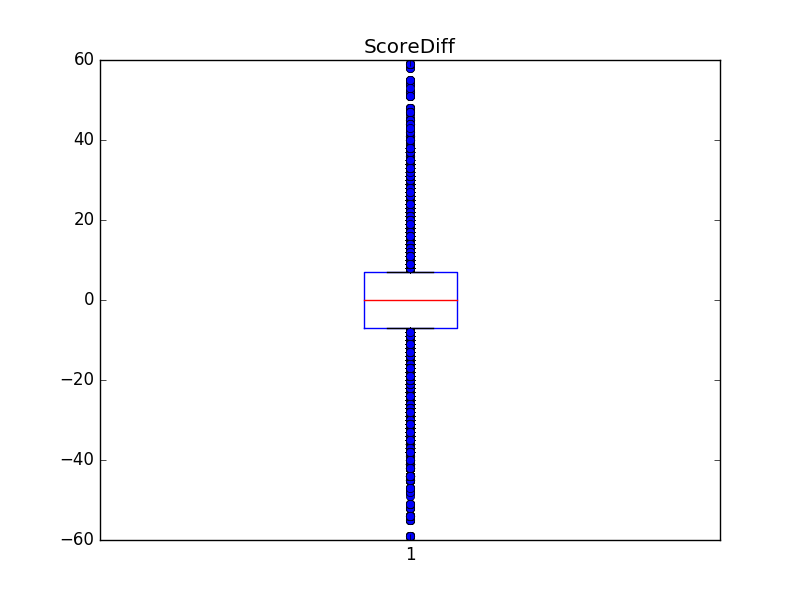
剔除缺失值的直方图 使用属性之间的相关关系填充缺失值的直方图



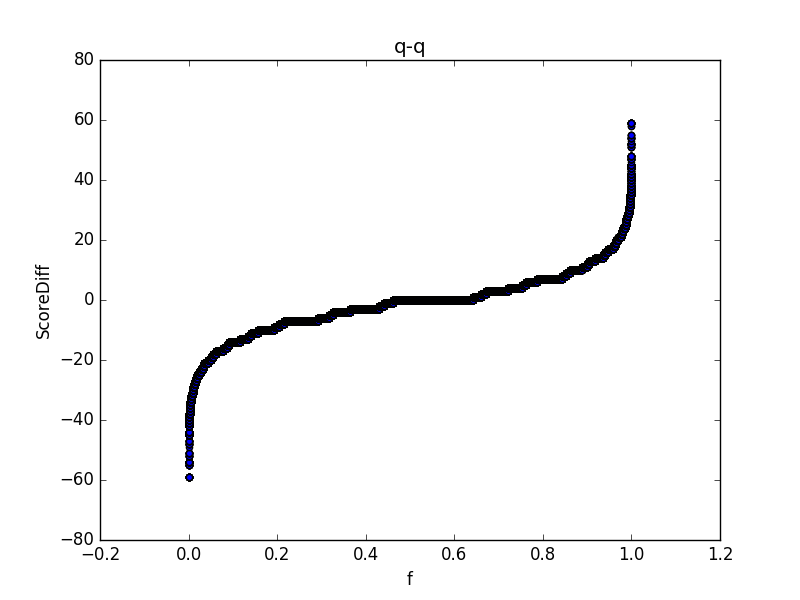
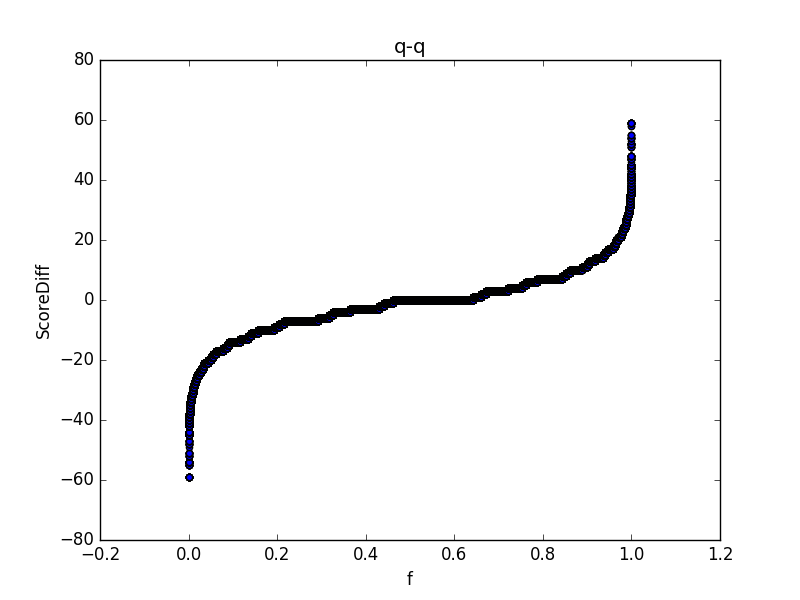
使用高频词填充



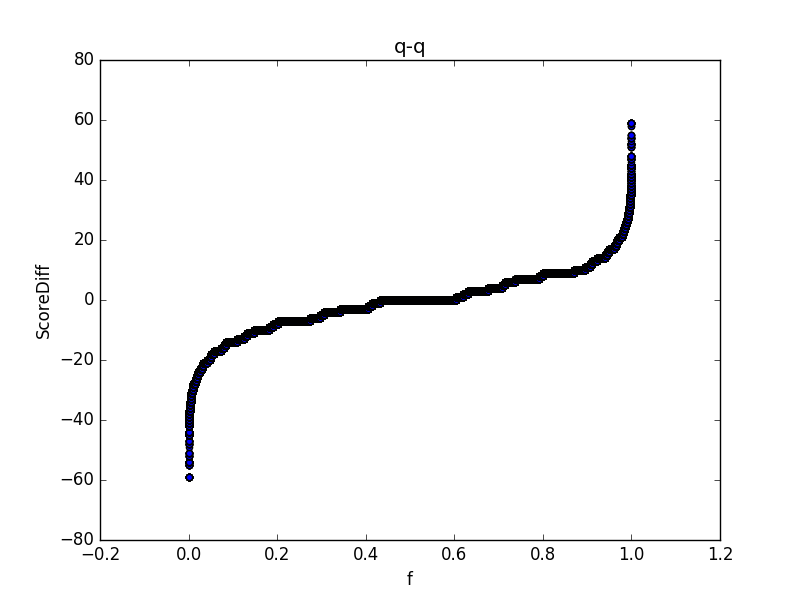
剔除缺失值的盒图 使用属性之间的相关关系填充缺失值的盒图



使用高频词填充

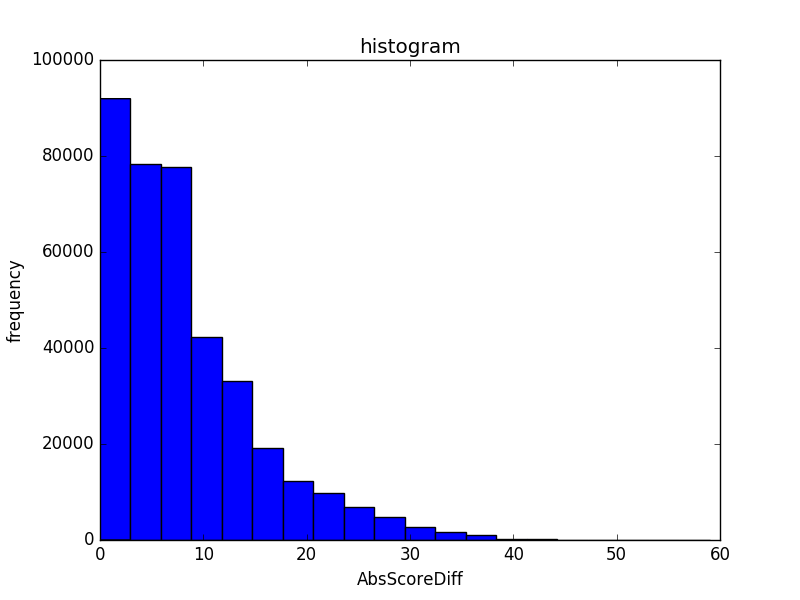
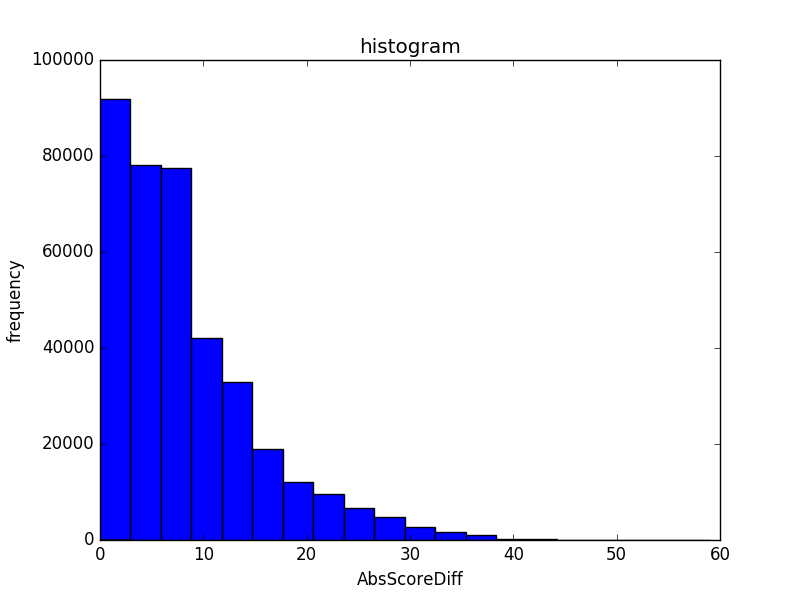


剔除缺失值的分位数图 使用属性之间的相关关系填充缺失值的分位数图

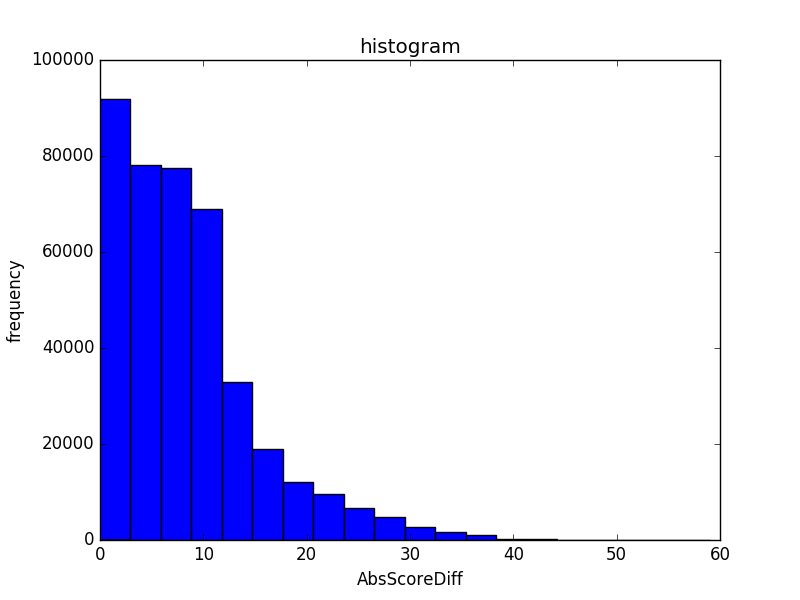


使用高频词填充

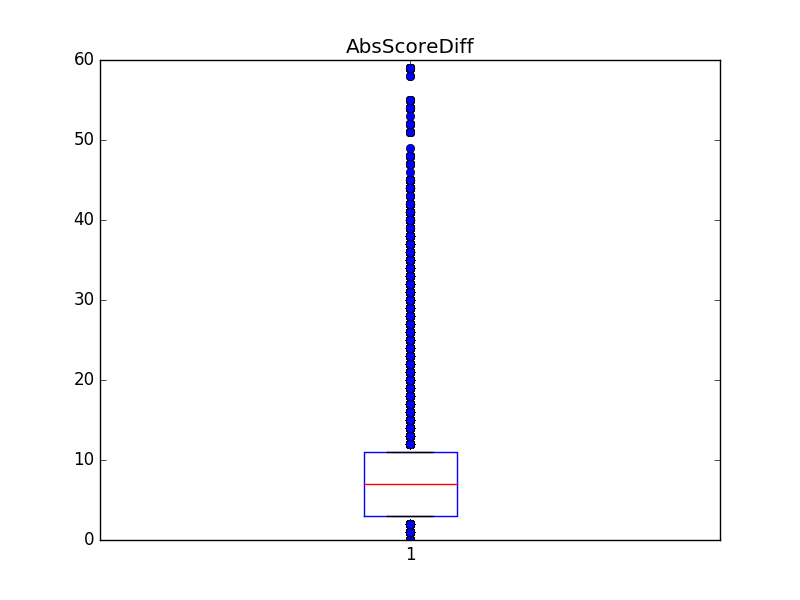
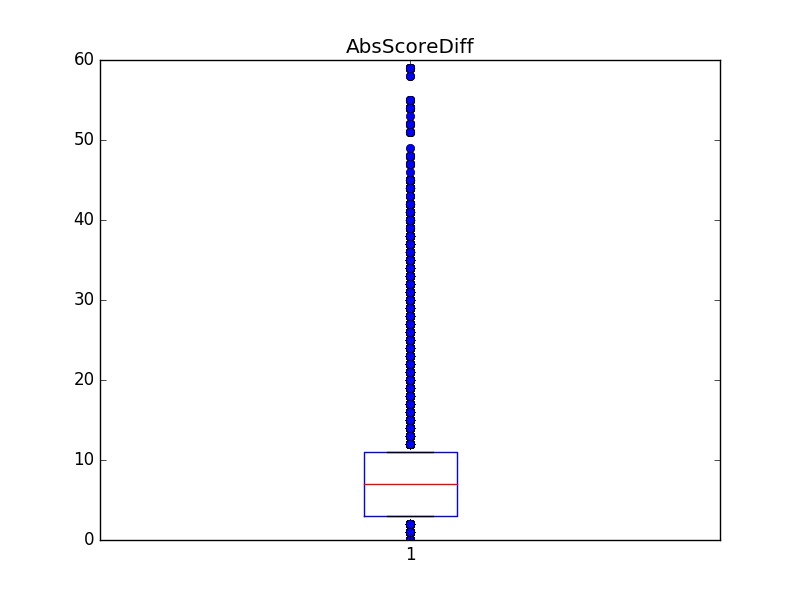
1. AbsScoreDiff



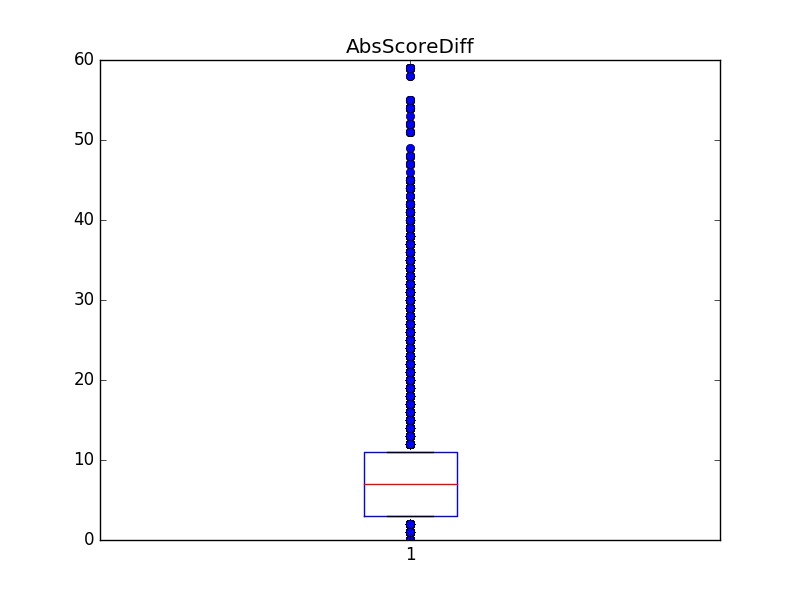
剔除缺失值 使用属性之间的相关关系



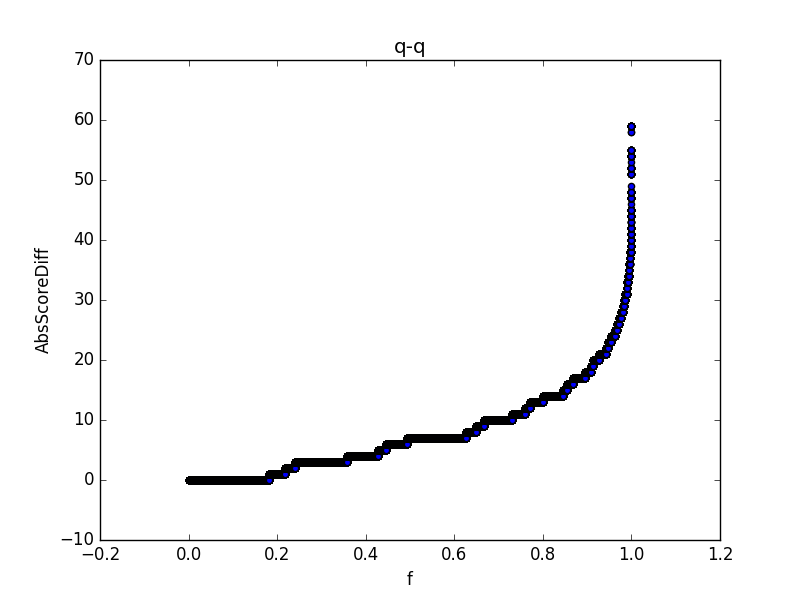
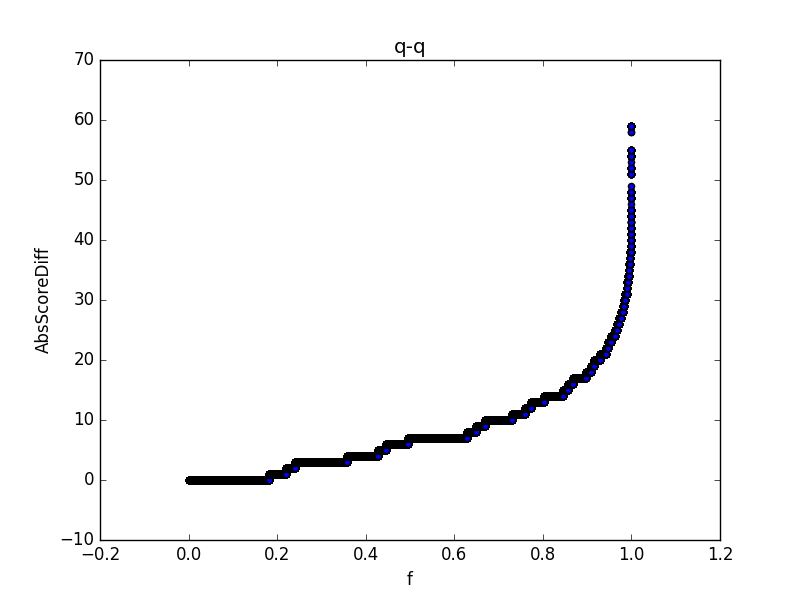
使用高频词填充



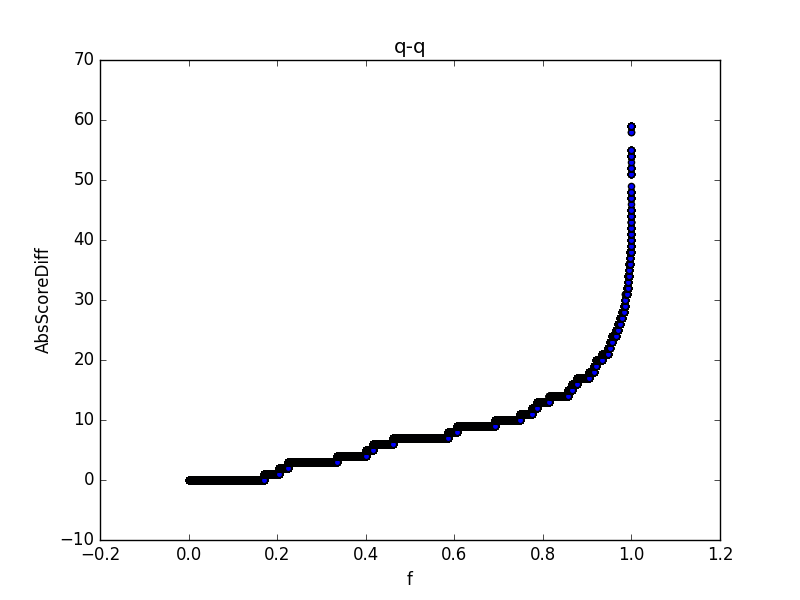
剔除缺失值 使用属性之间的相关关系



使用高频词填充

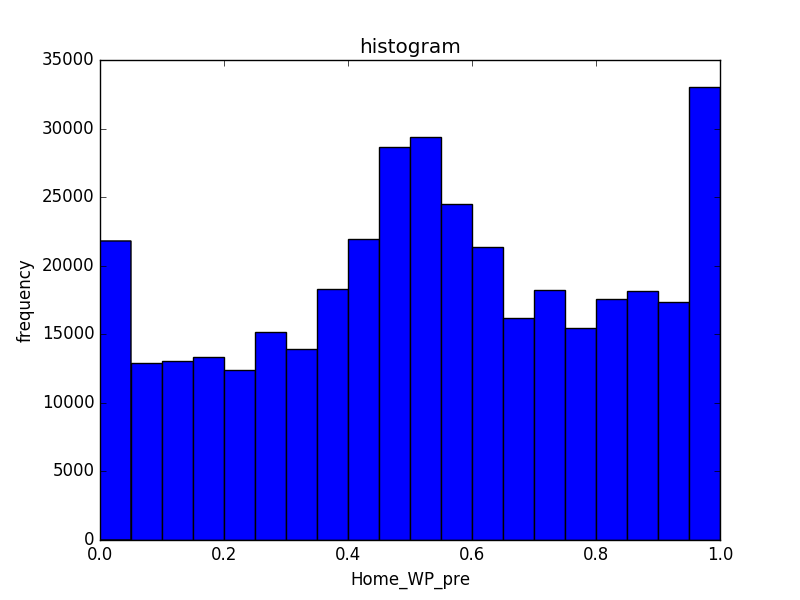
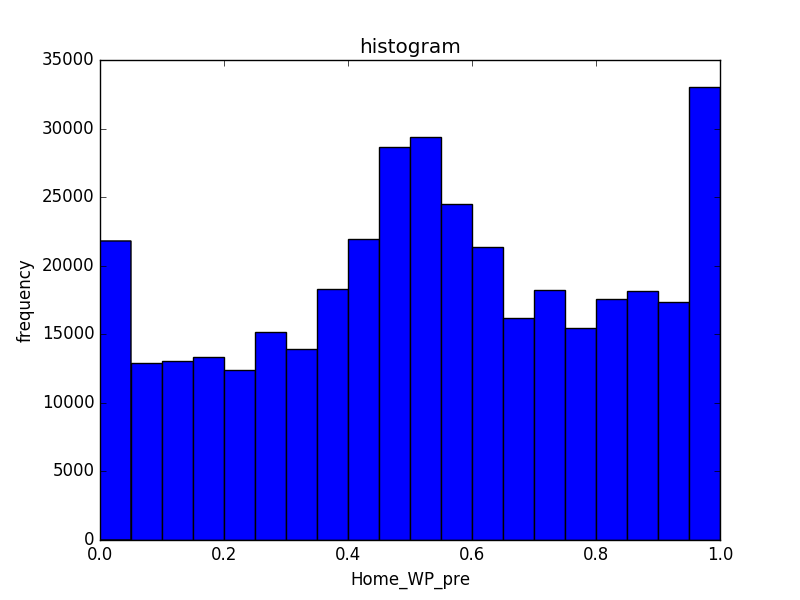


剔除缺失值 使用属性之间的相关关系

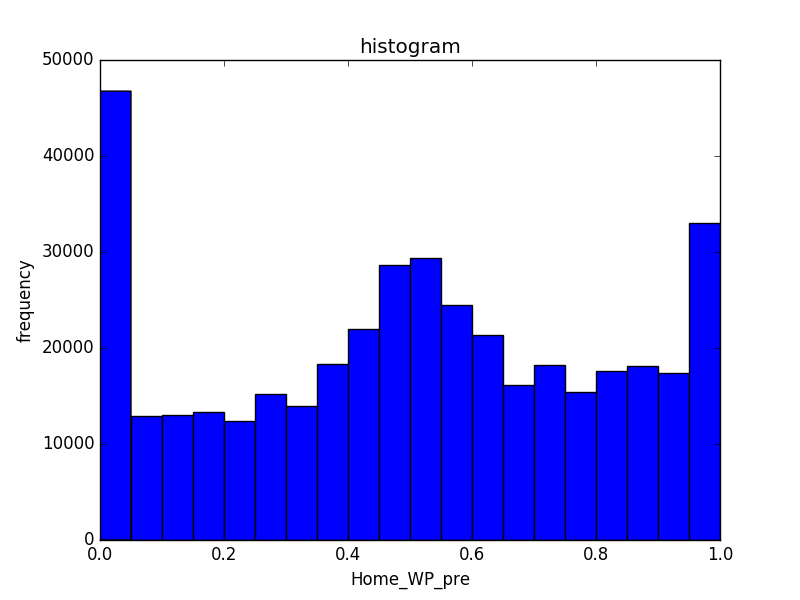


使用高频词填充

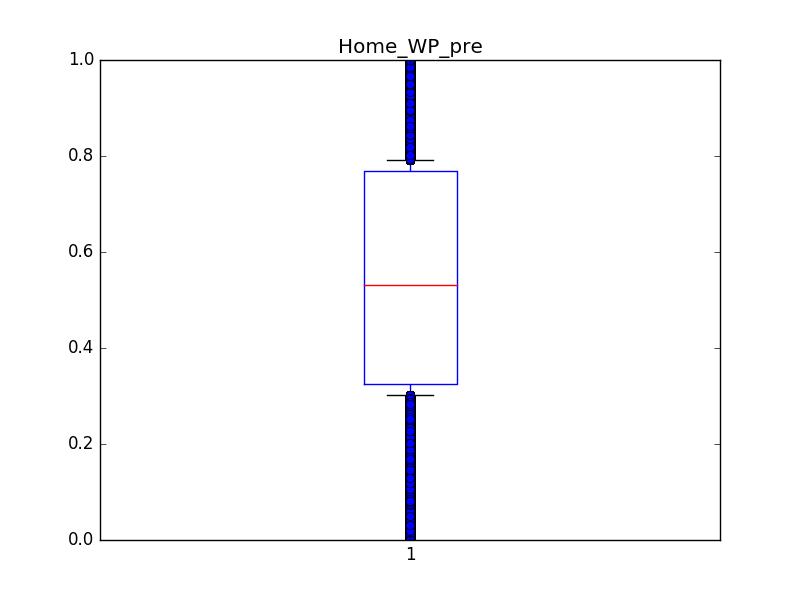
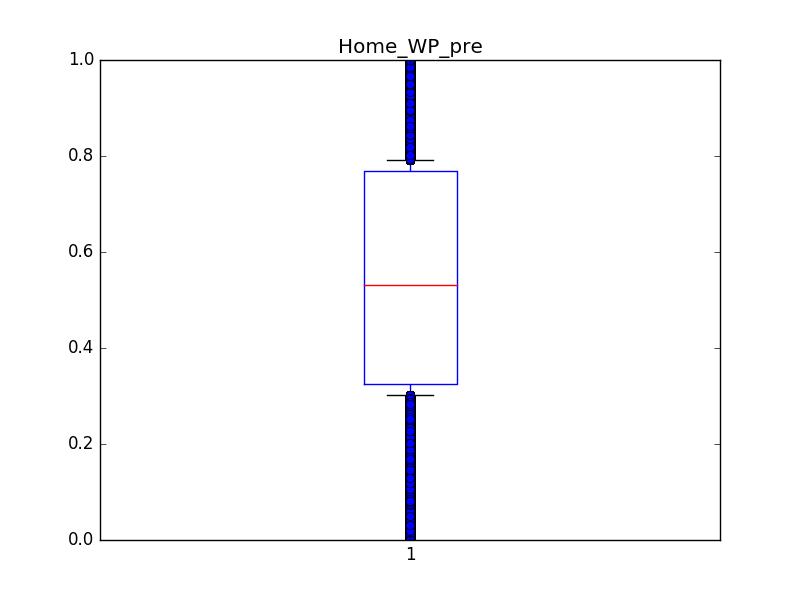
1. Home\_WP\_pre



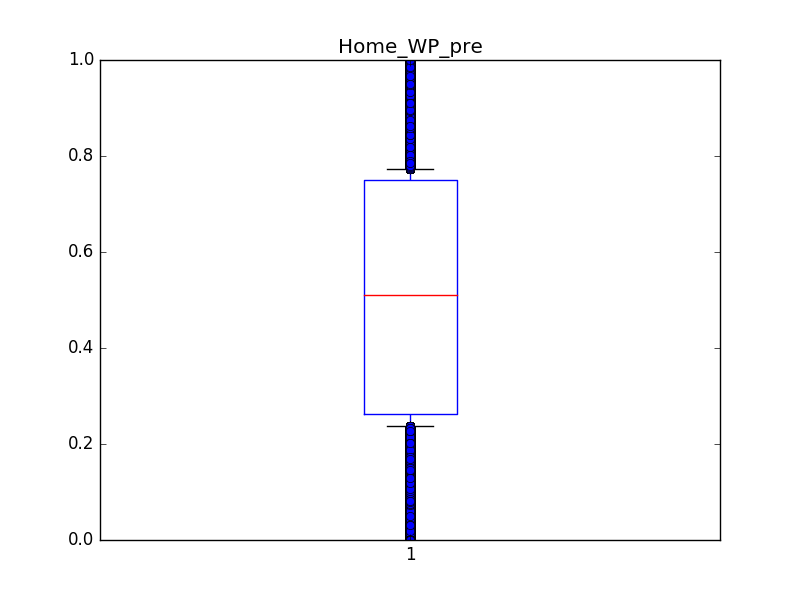
剔除缺失值 使用属性之间的相关关系



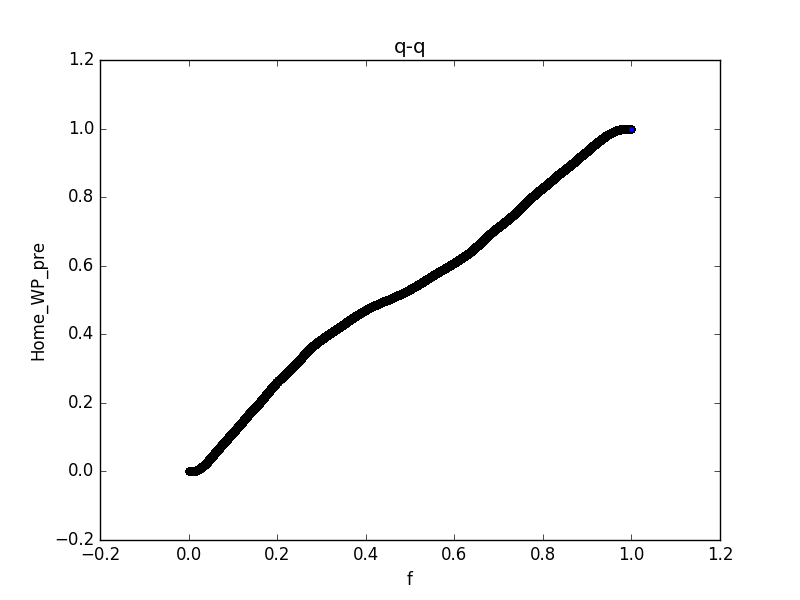
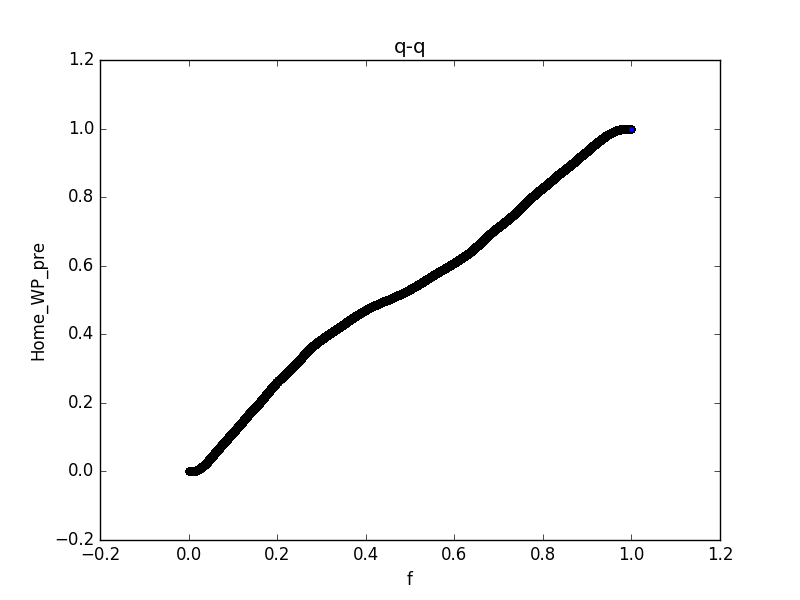
使用高频词填充



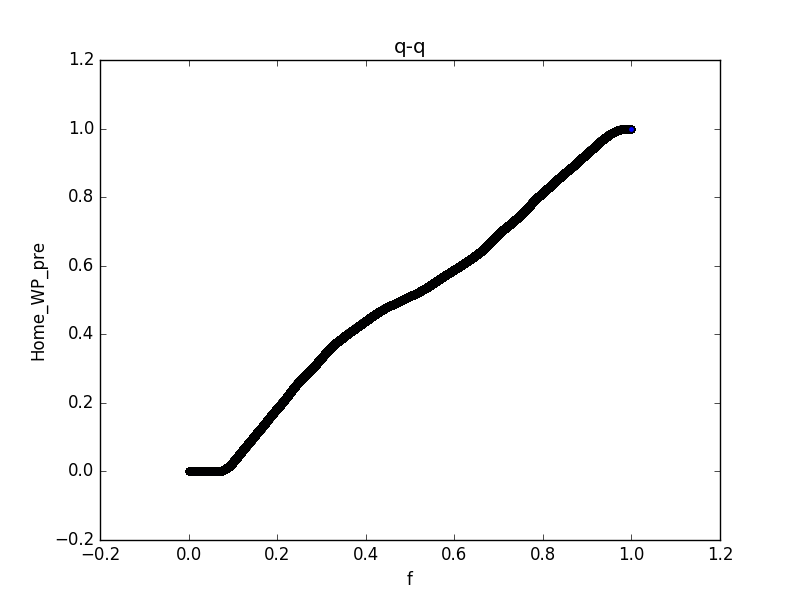
剔除缺失值 使用属性之间的相关关系



使用高频词填充

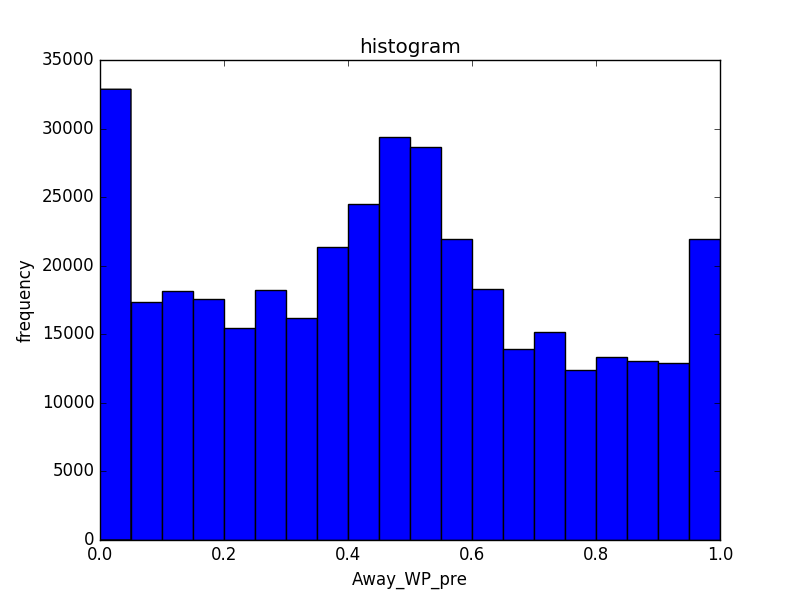
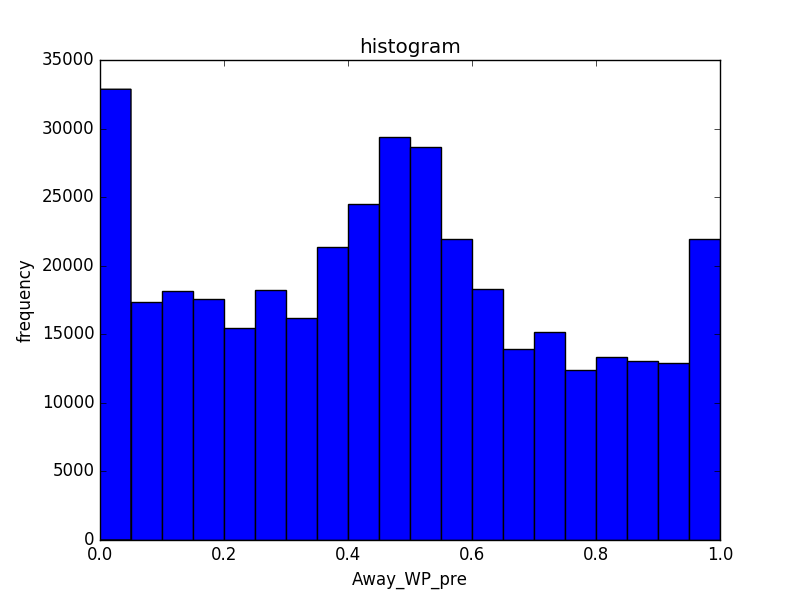


剔除缺失值 使用属性之间的相关关系

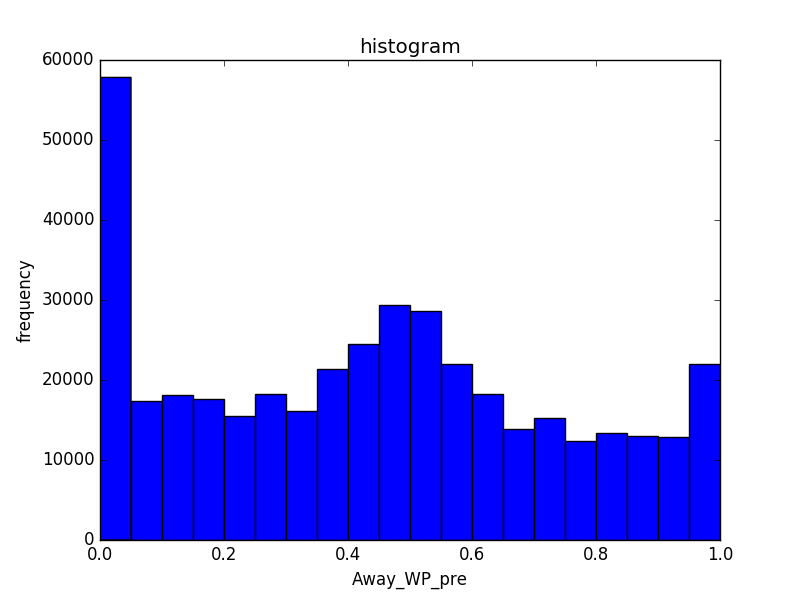


使用高频词填充

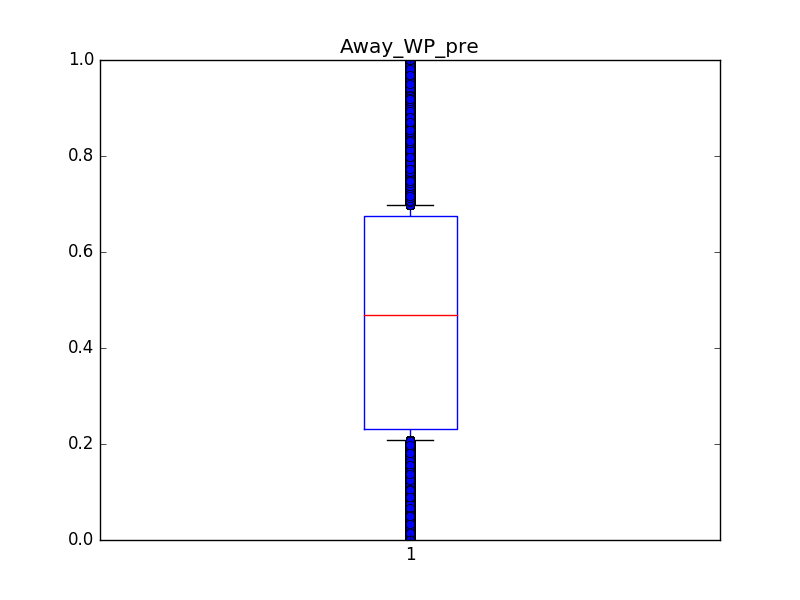
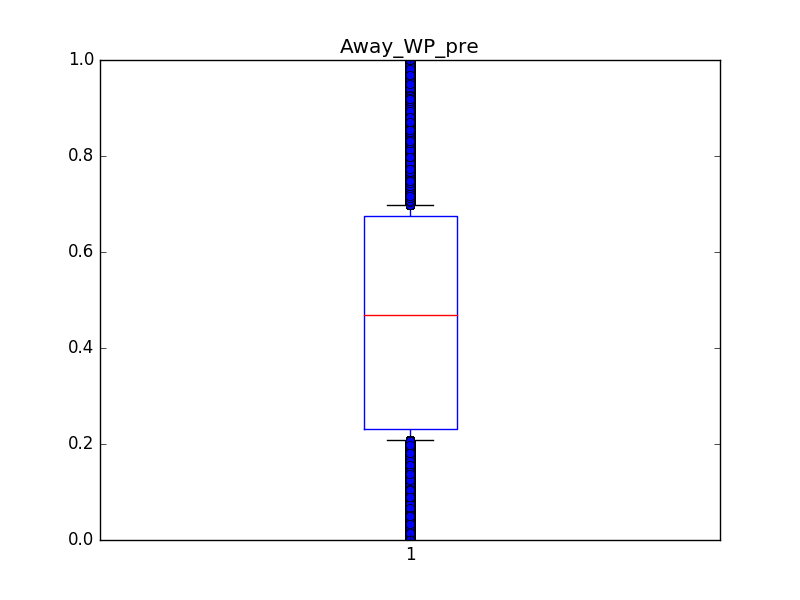
1. Away\_WP\_pre



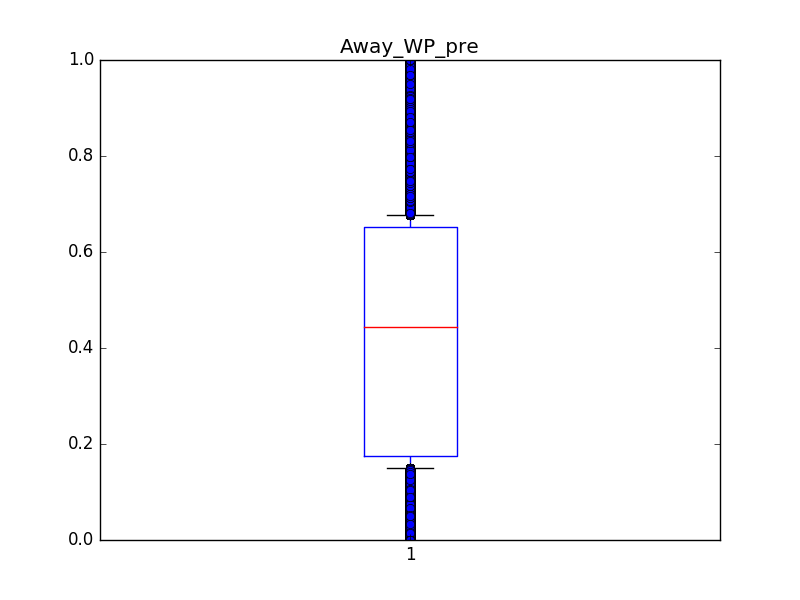
剔除缺失值 使用属性之间的相关关系



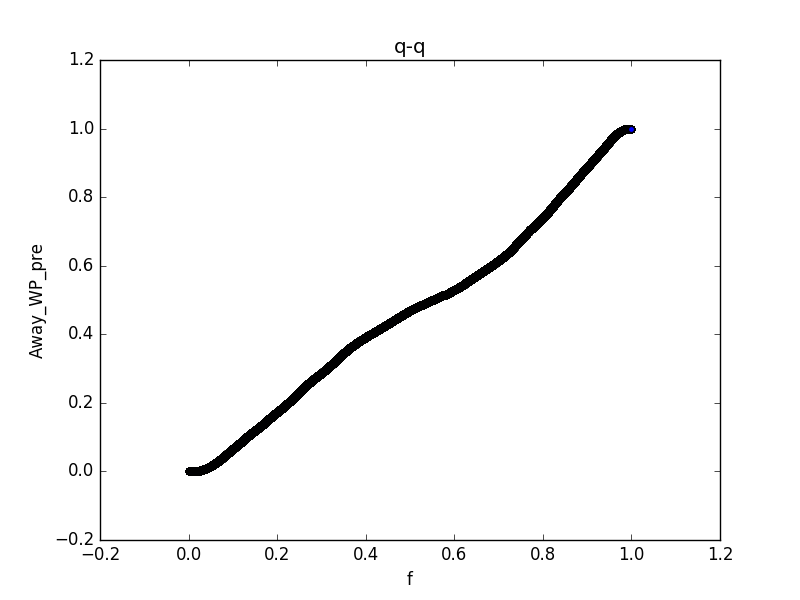
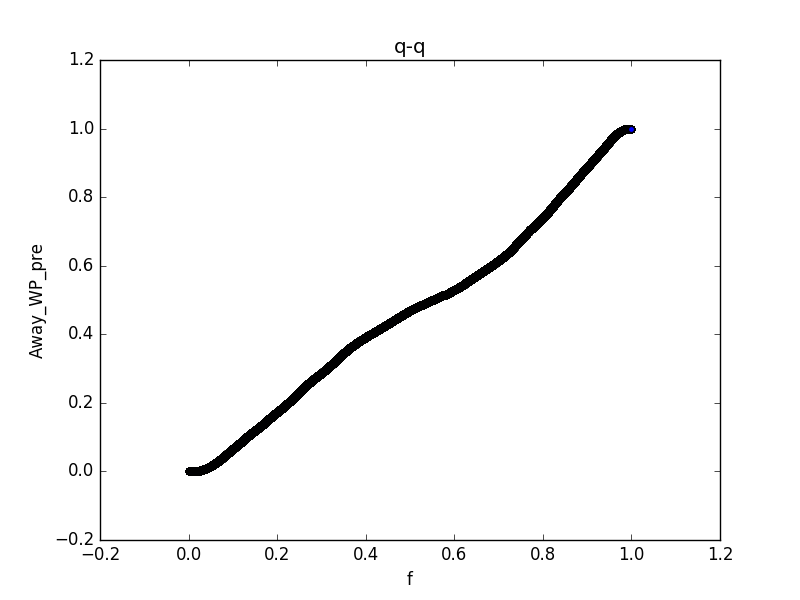
使用高频词填充



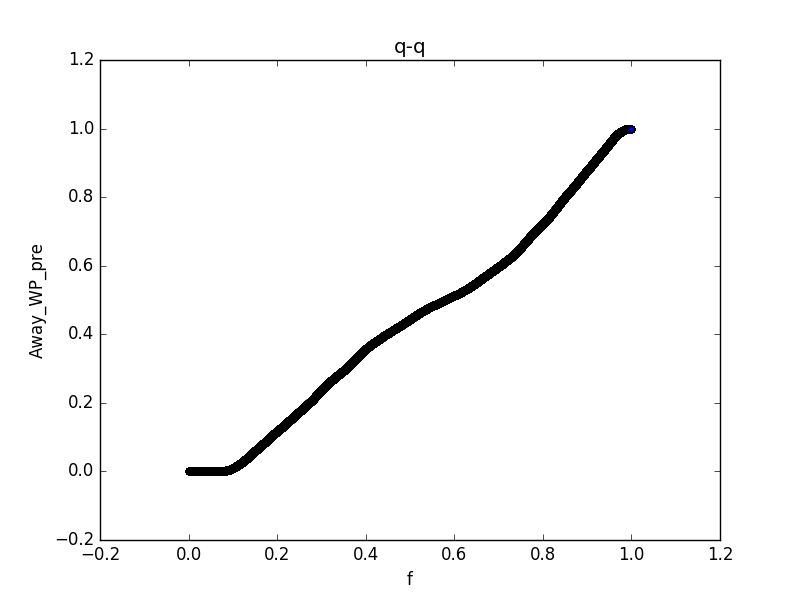
剔除缺失值 使用属性之间的相关关系



使用高频词填充

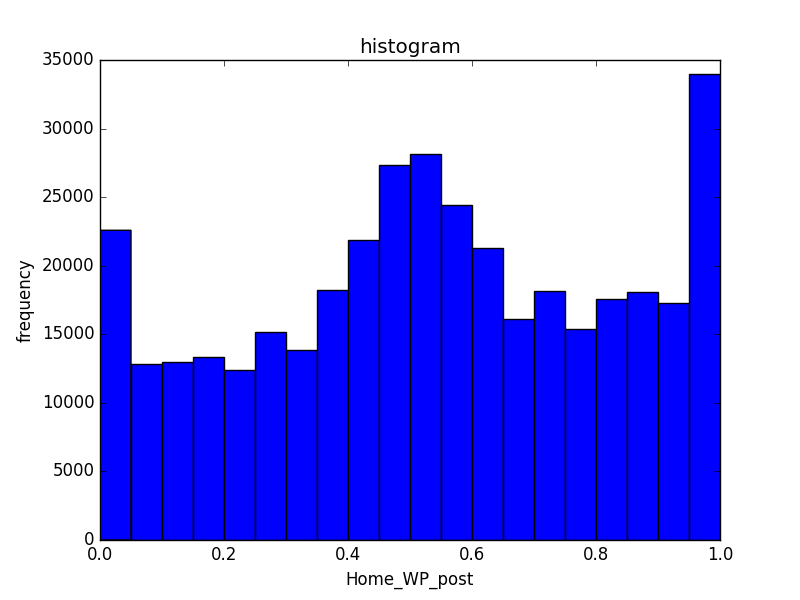
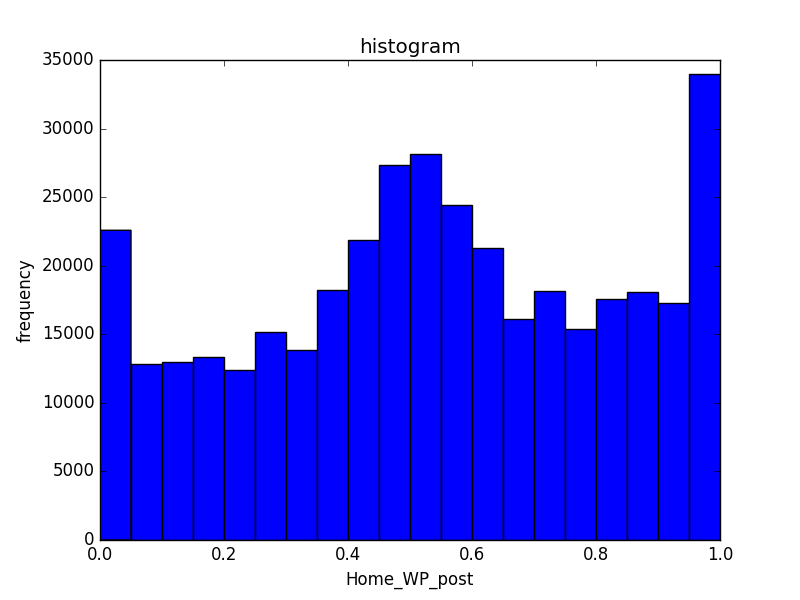


剔除缺失值 使用属性之间的相关关系

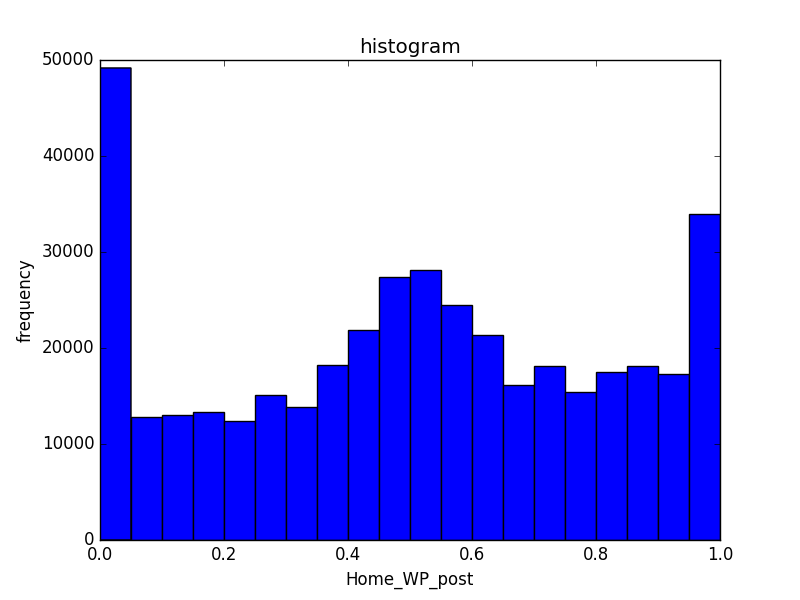


使用高频词填充

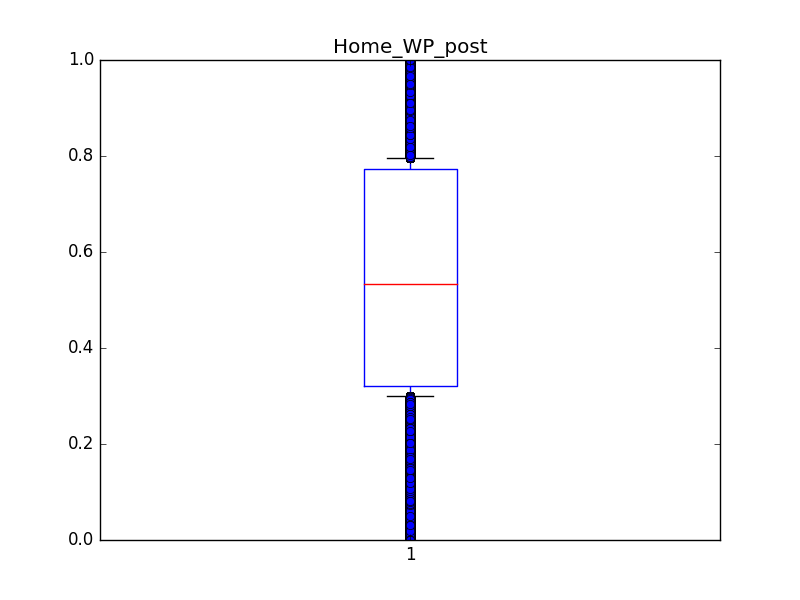
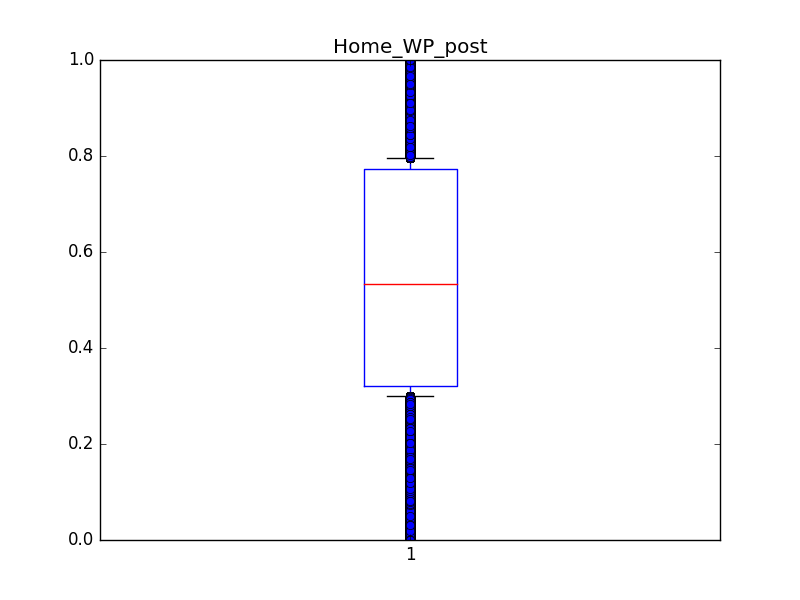
1. Home\_WP\_post



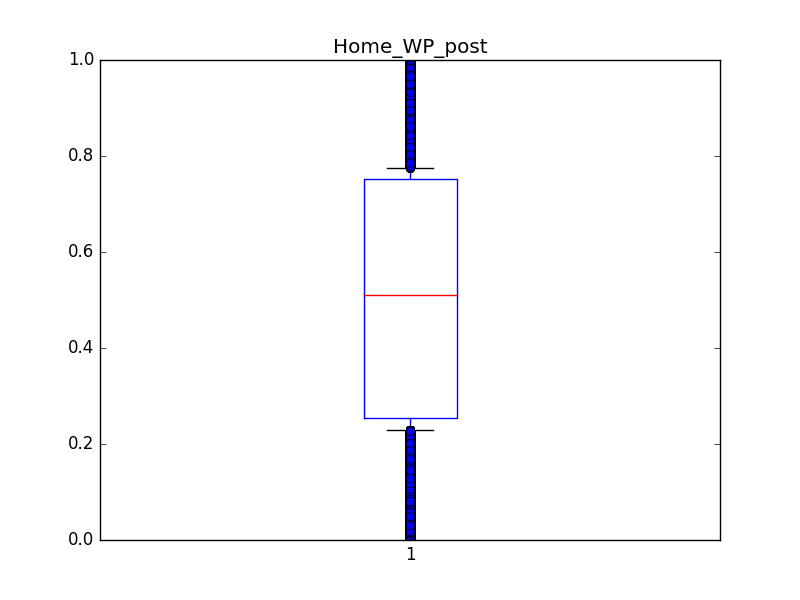
剔除缺失值 使用属性之间的相关关系填充



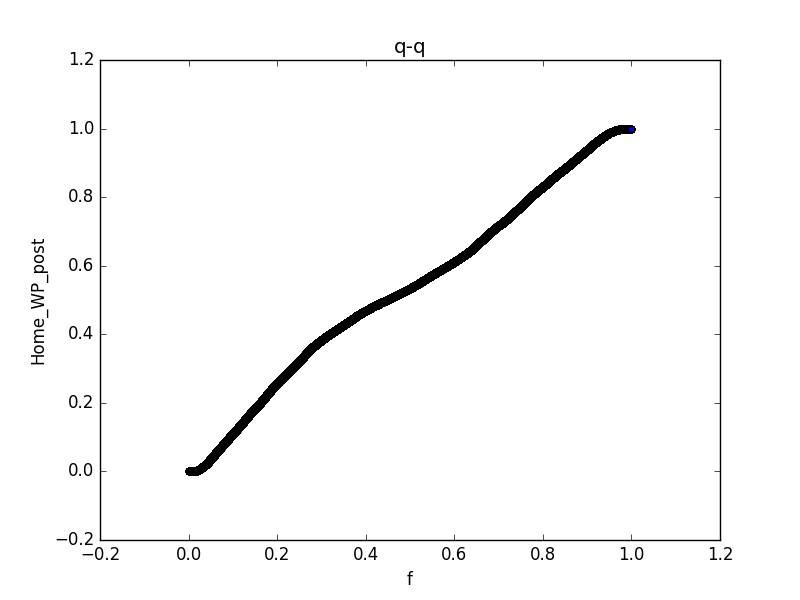
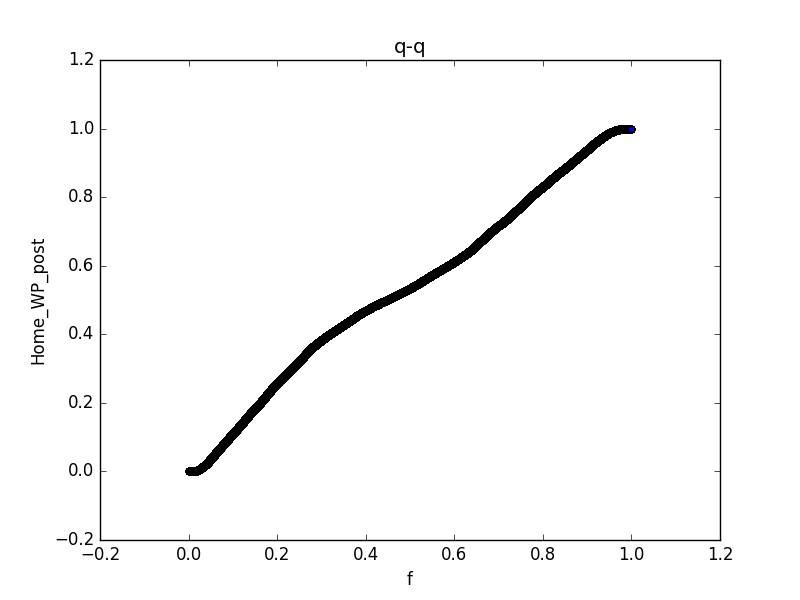
使用高频词填充



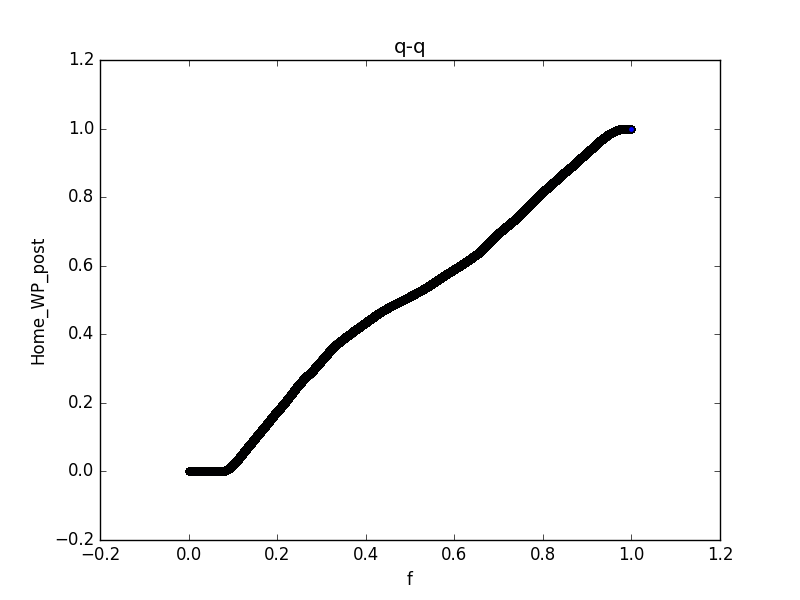
剔除缺失值 使用属性之间的相关关系填充



使用高频词填充

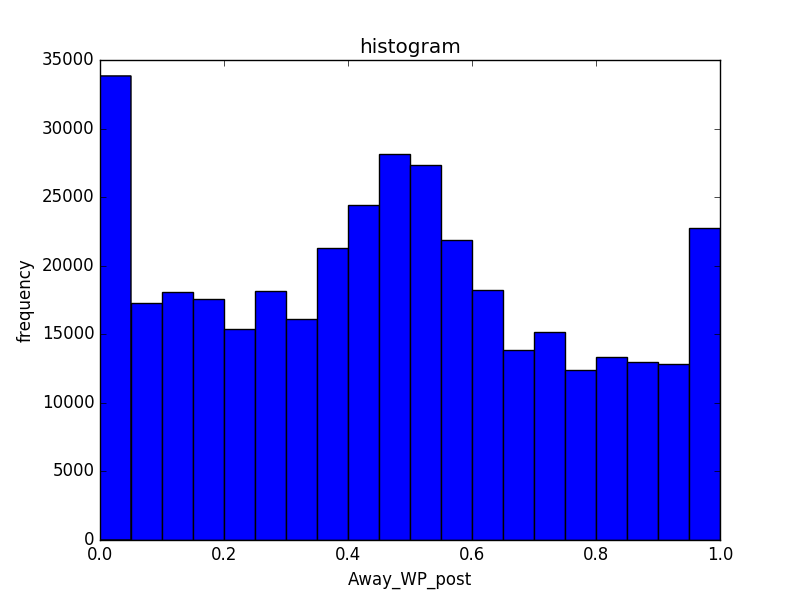
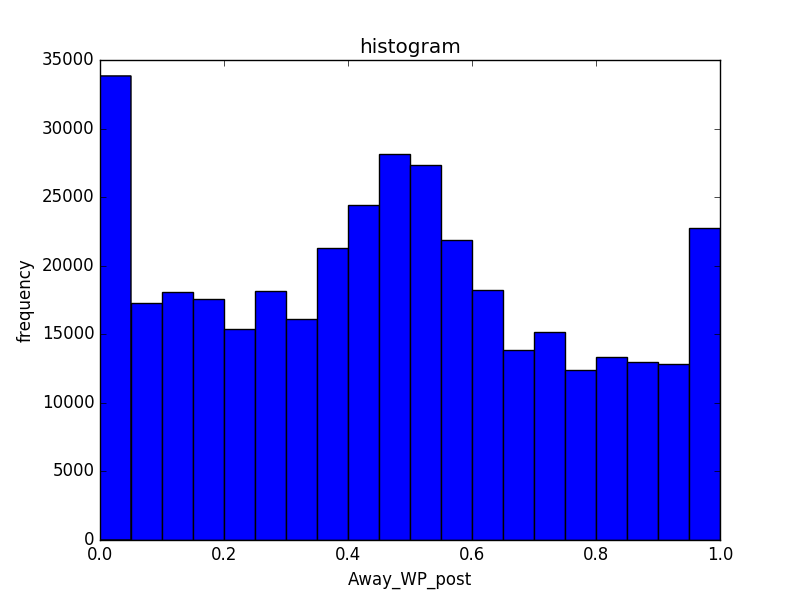


剔除缺失值 使用属性之间的相关关系填充

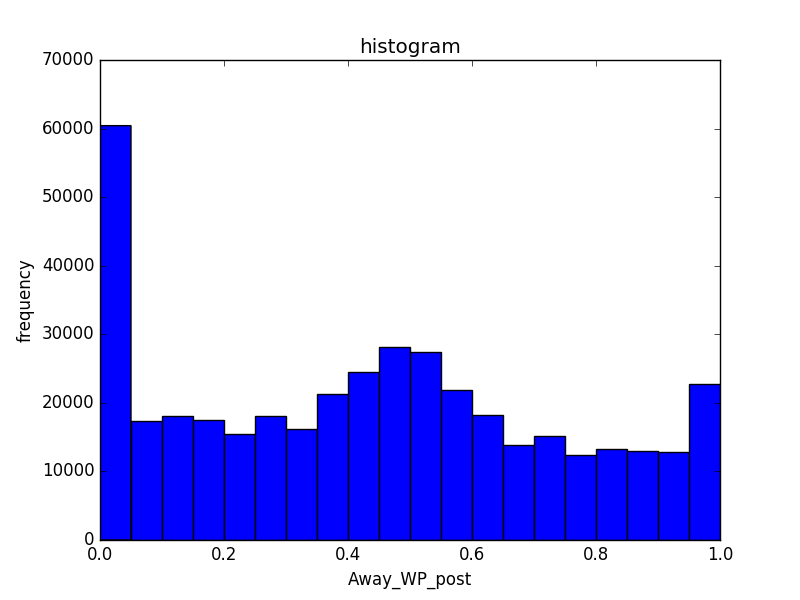


使用高频词填充

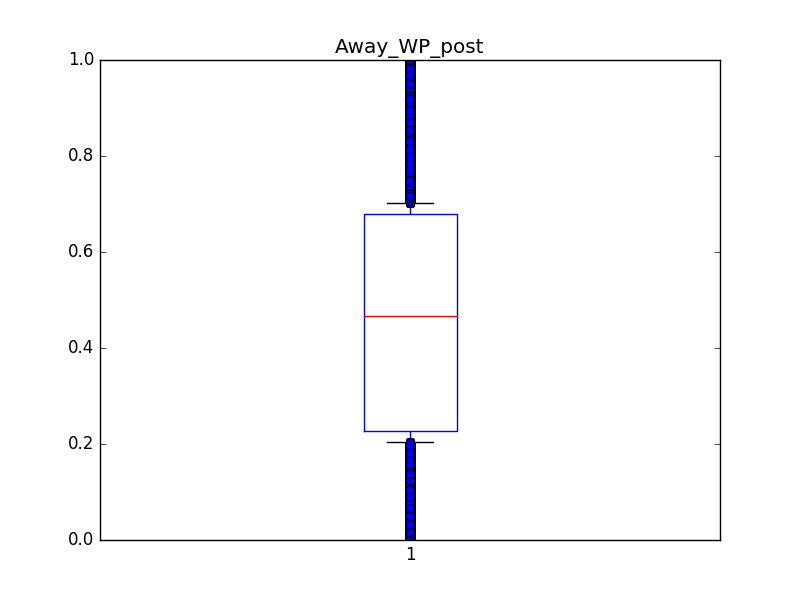
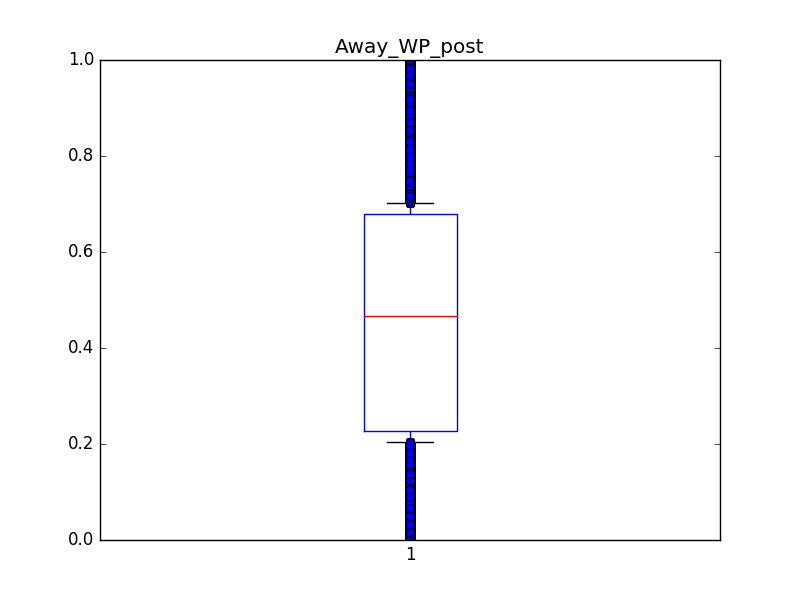
1. Away\_WP\_post



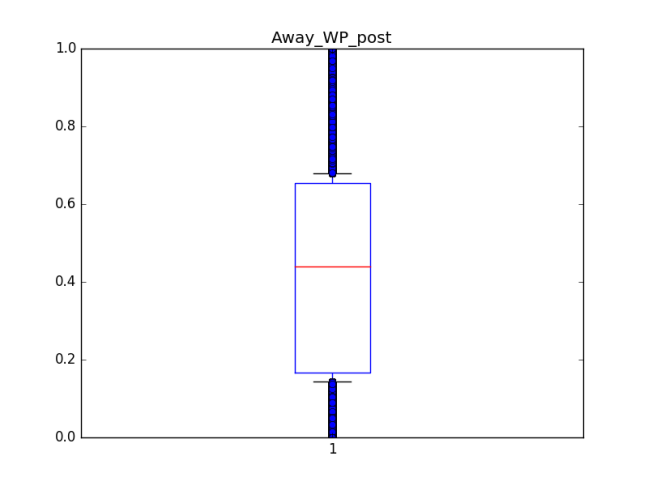
剔除缺失值 使用属性之间的相关关系填充



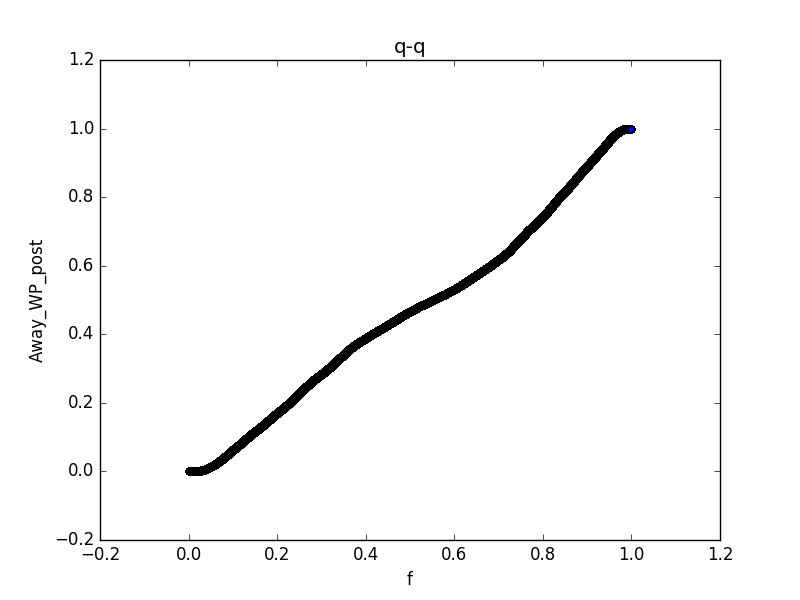
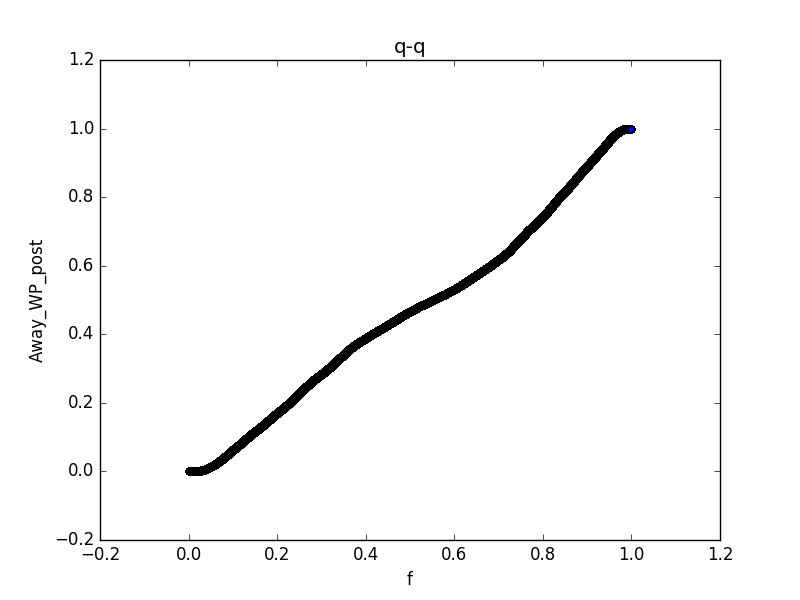
使用高频词填充



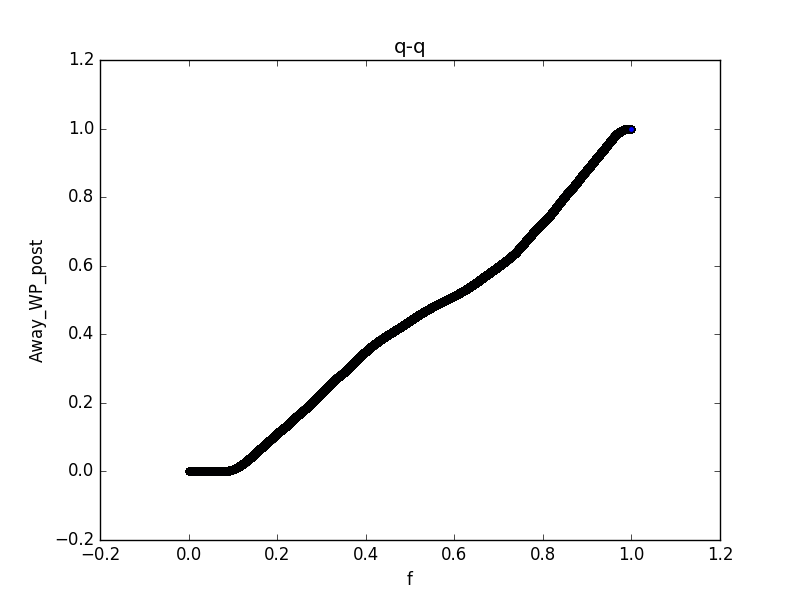
剔除缺失值 使用属性之间的相关关系



使用高频词填充

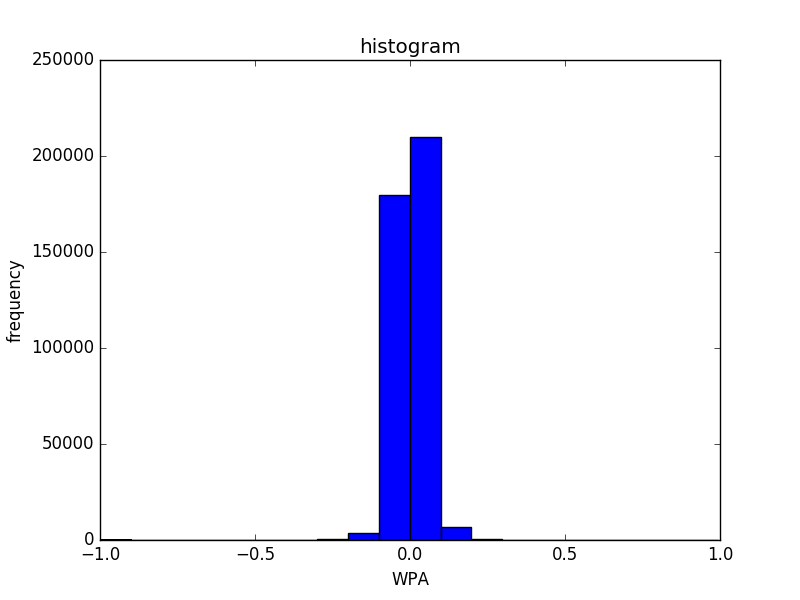
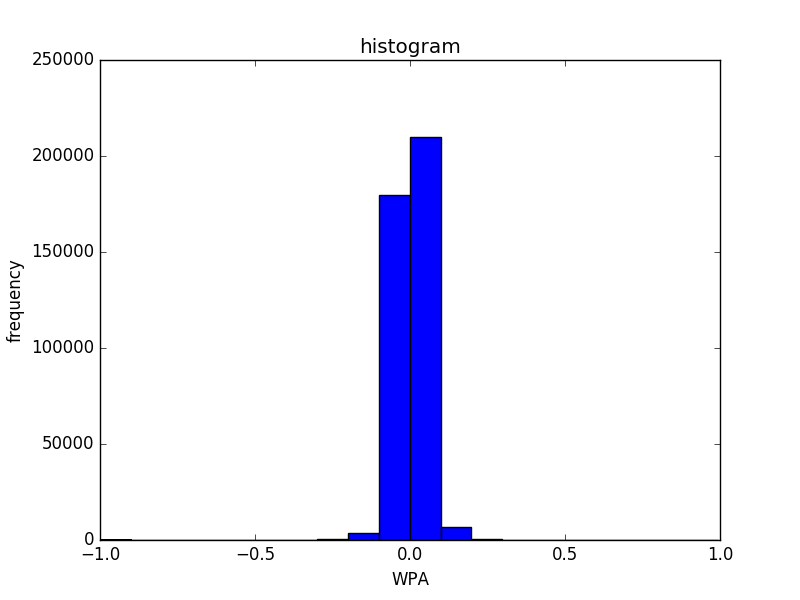


剔除缺失值 使用属性之间的相关关系

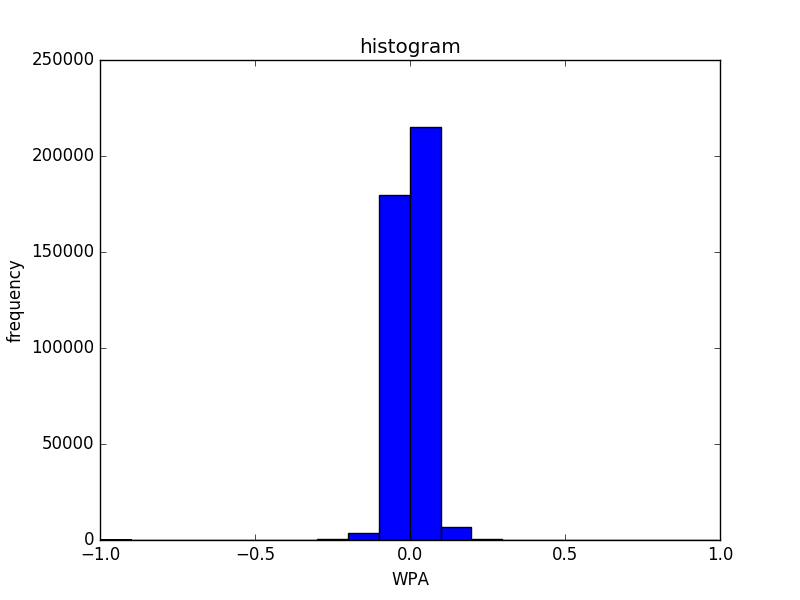


使用高频词填充

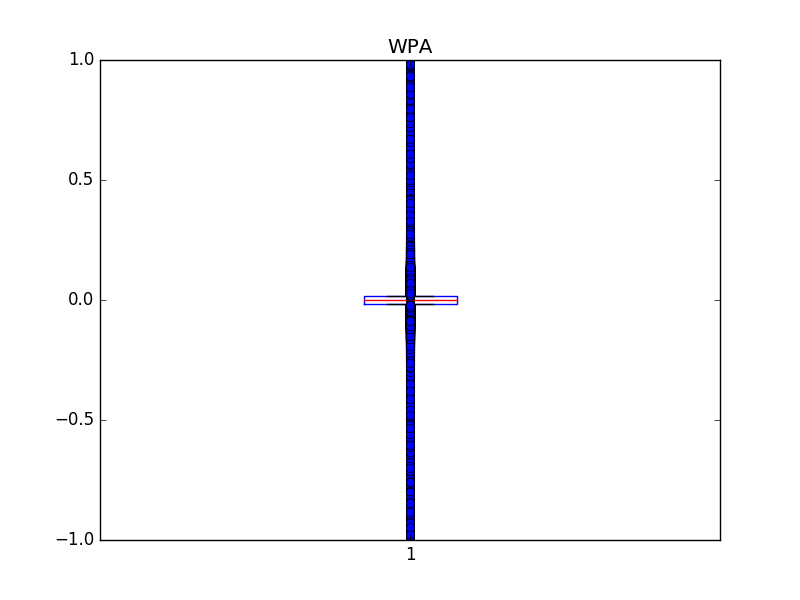
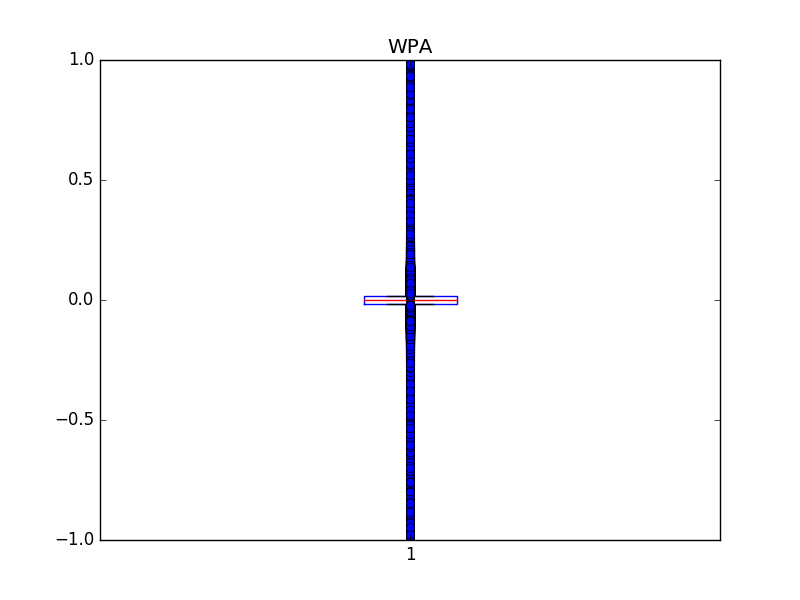
（7）WPA



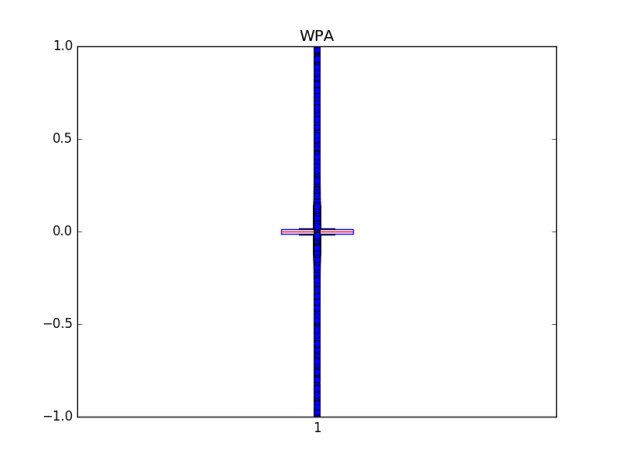
剔除缺失值 使用属性之间的相关关系



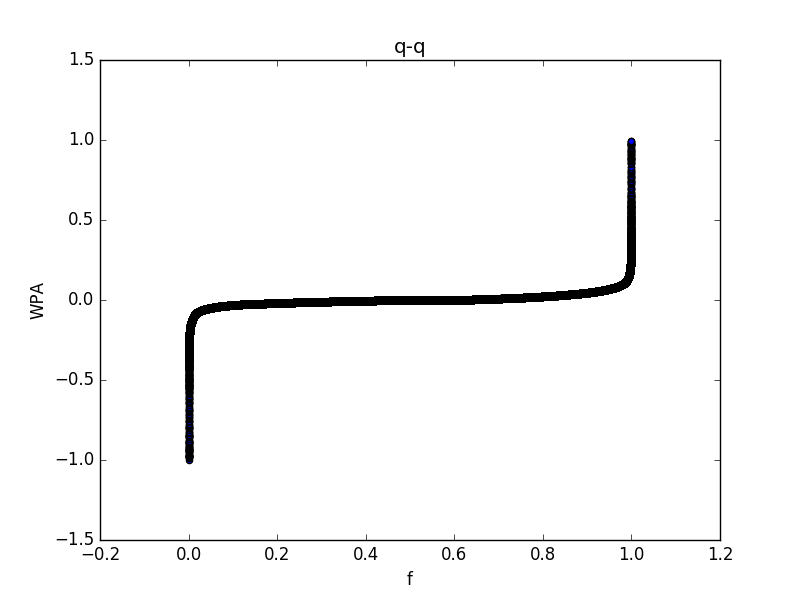
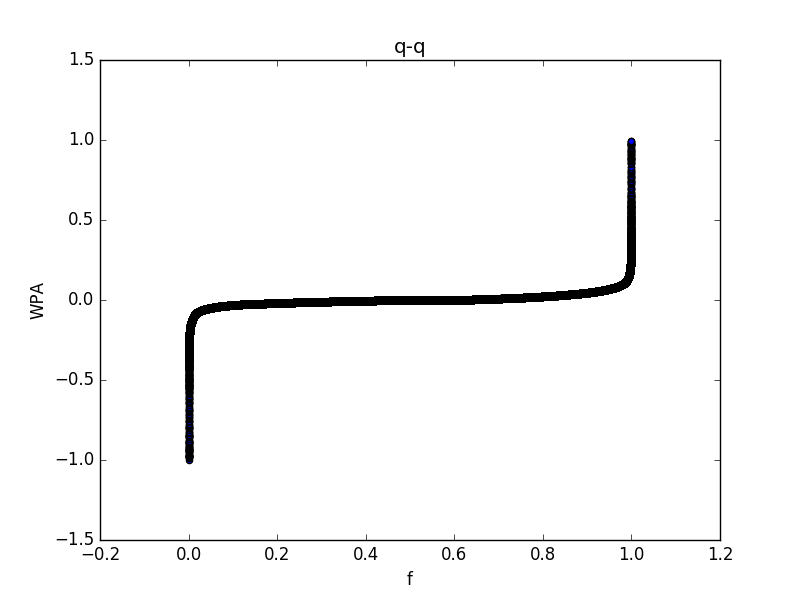
使用高频词填充



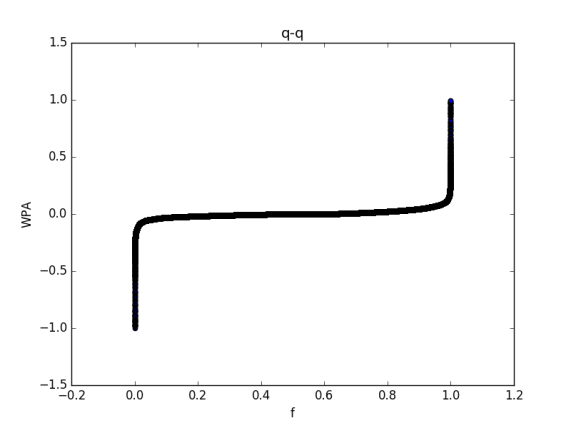
剔除缺失值 使用属性之间的相关关系



使用高频词填充

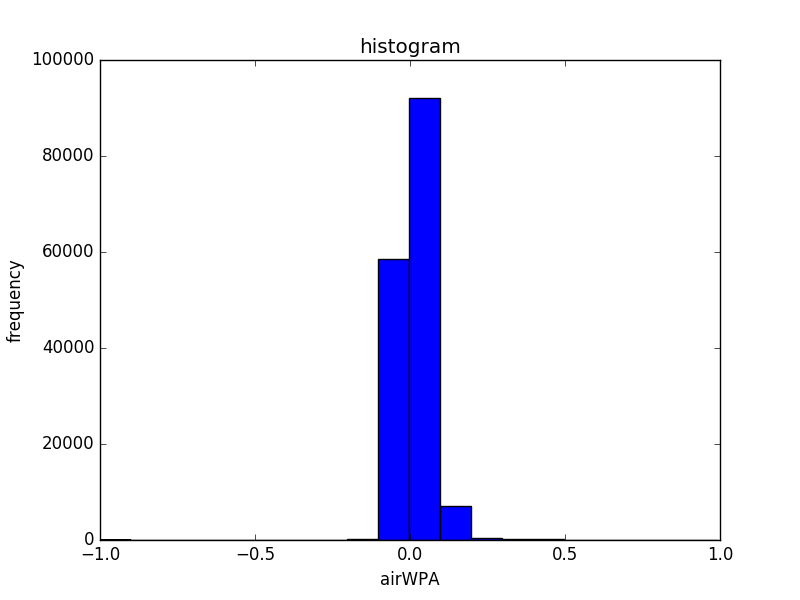
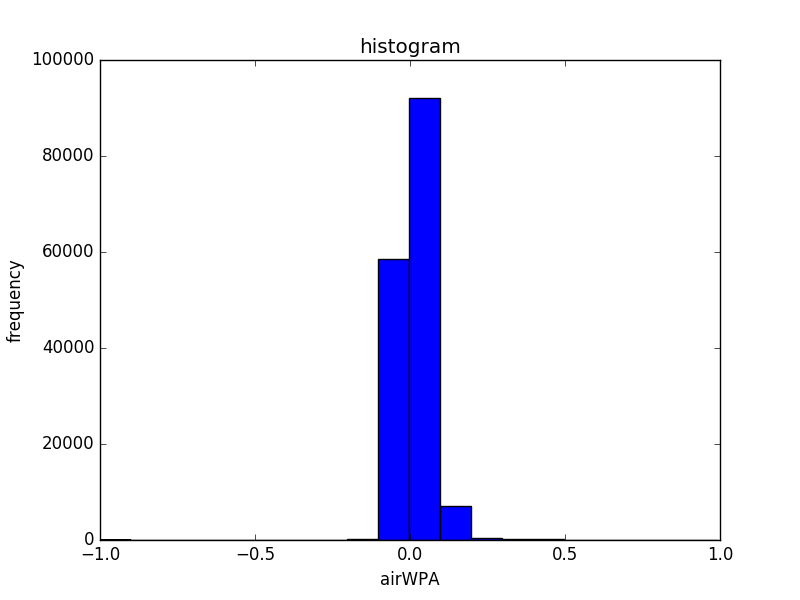


剔除缺失值 使用属性之间的相关关系

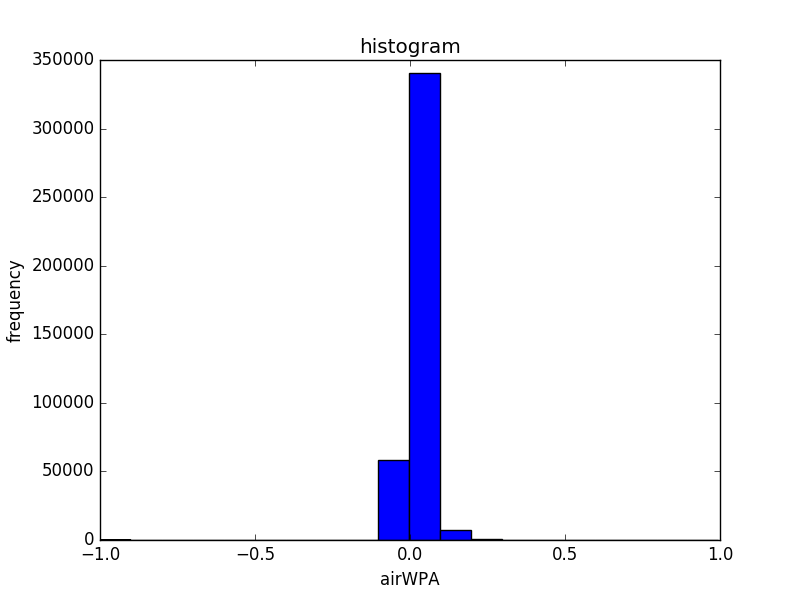


使用高频词填充

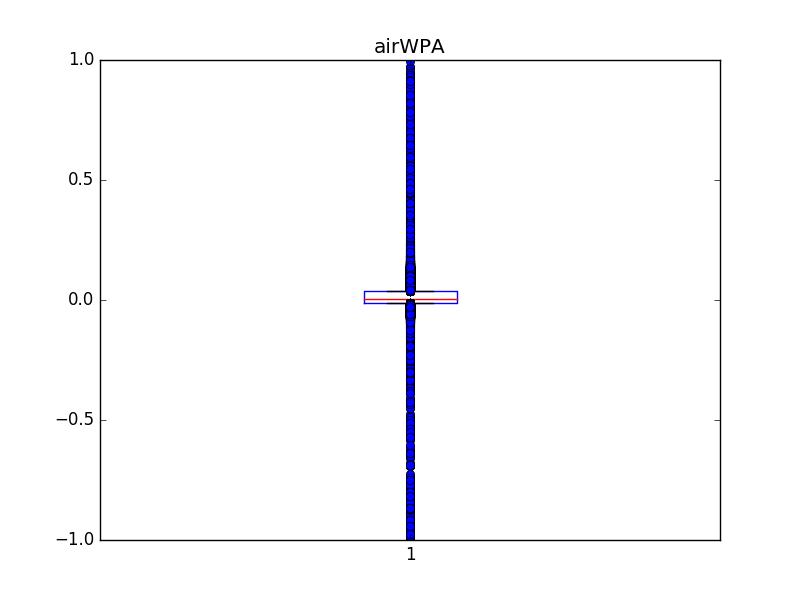
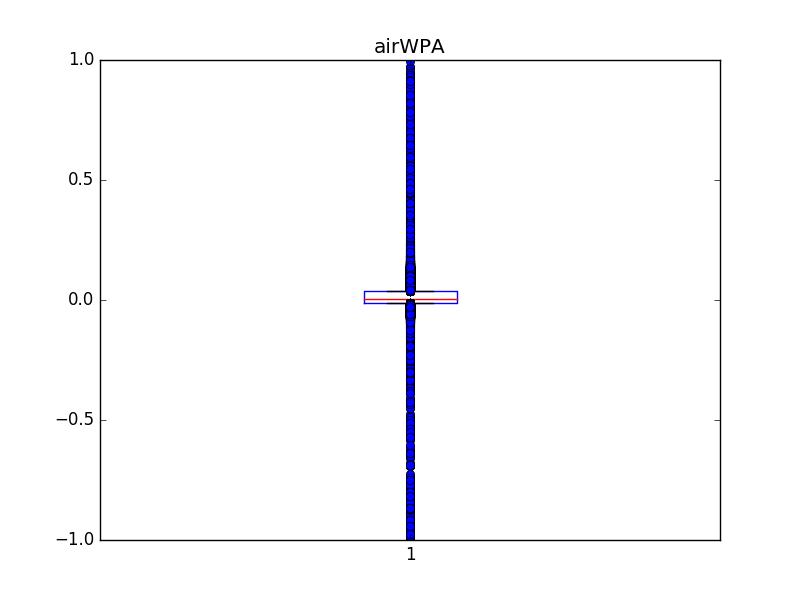
（8）airWPA



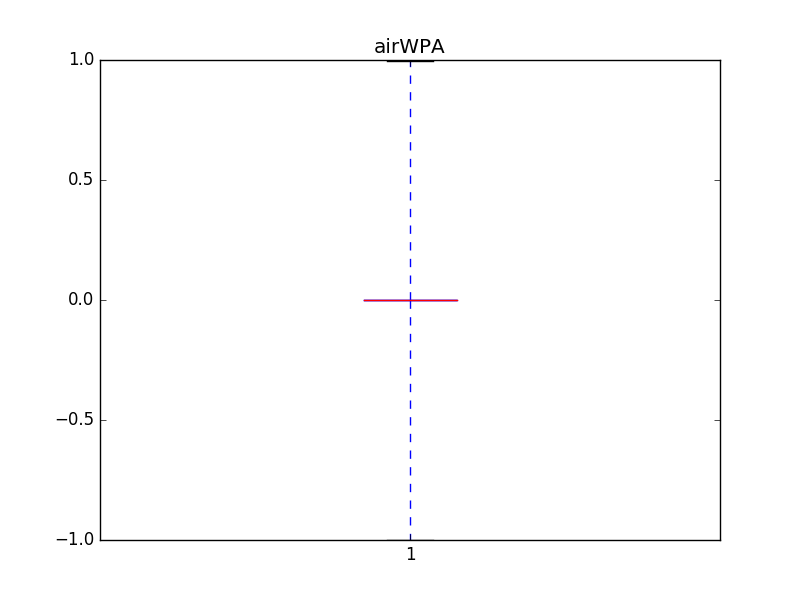
剔除缺失值 使用属性之间的相关关系填充



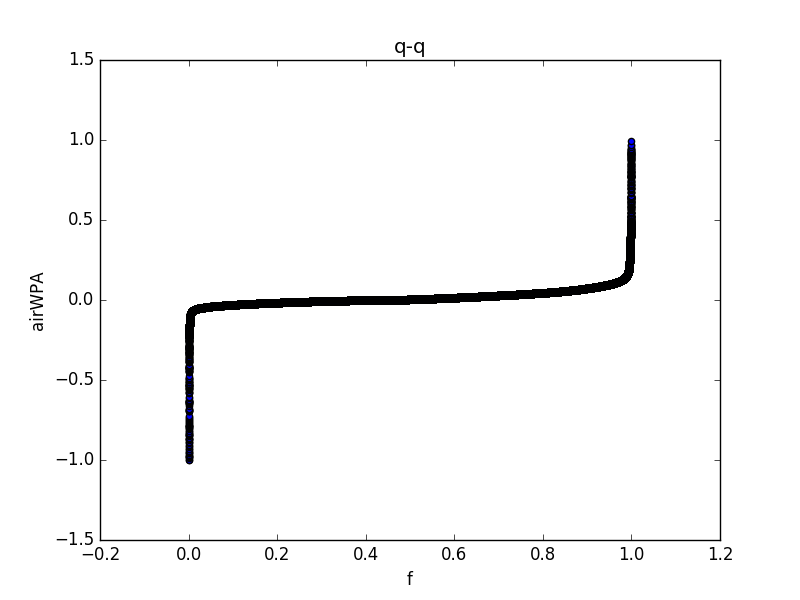
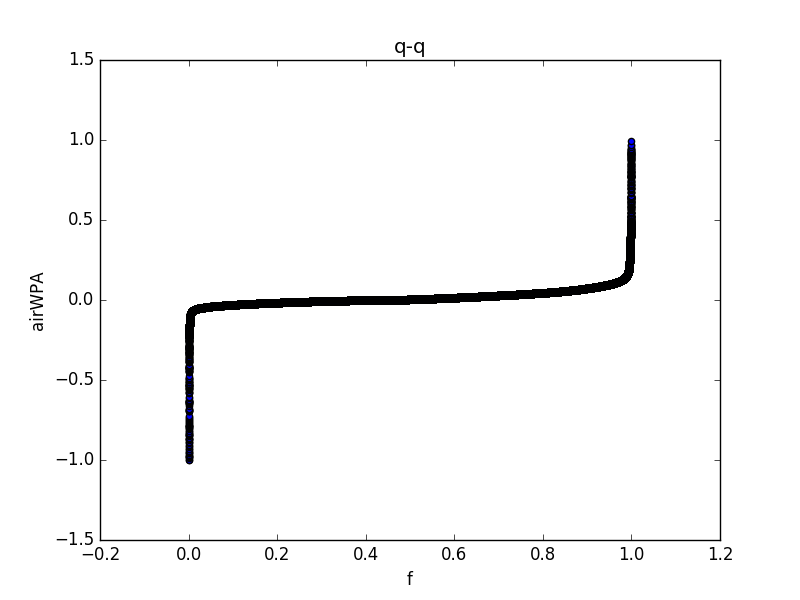
使用高频词填充



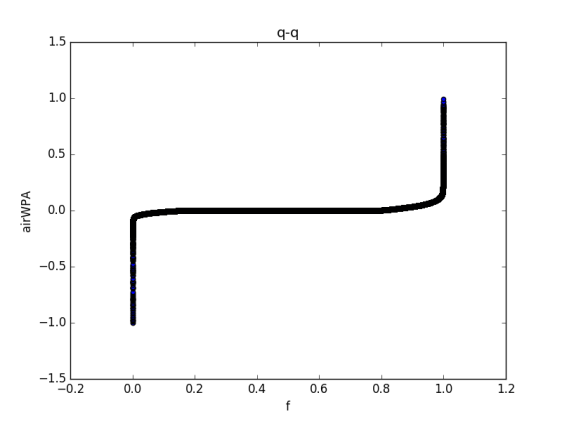
剔除缺失值 使用属性之间的相关关系填充



使用高频词填充

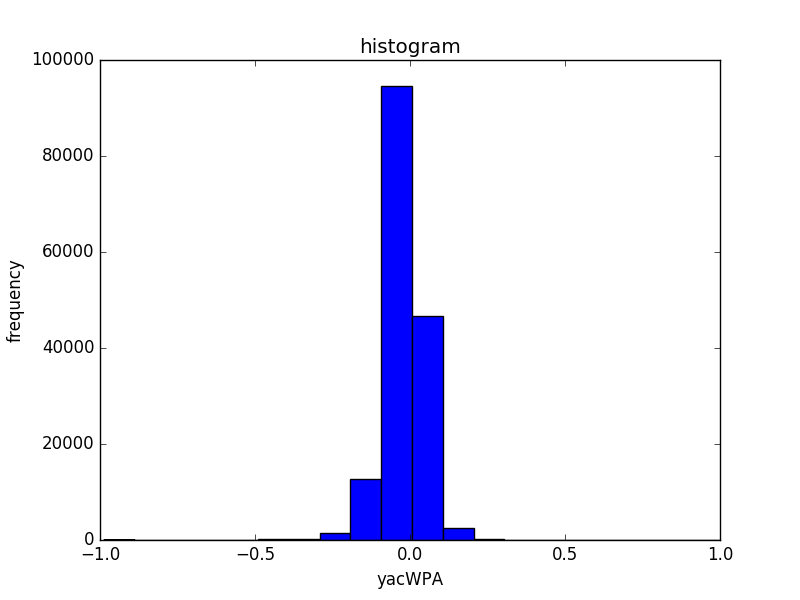
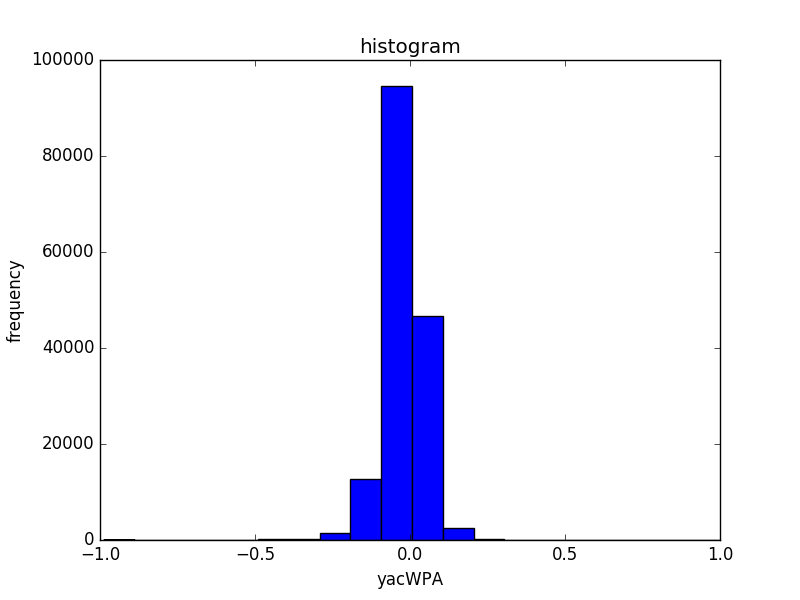


剔除缺失值 使用属性之间的相关关系填充

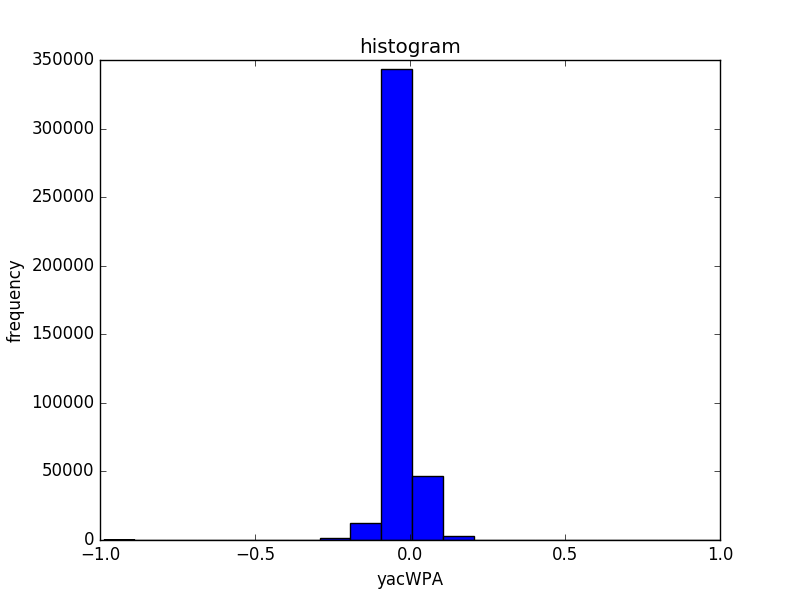


使用高频词填充

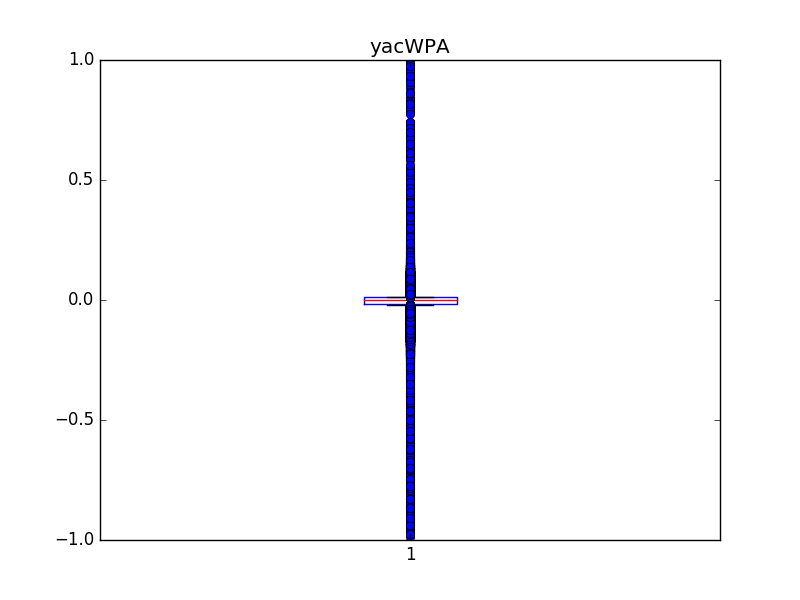
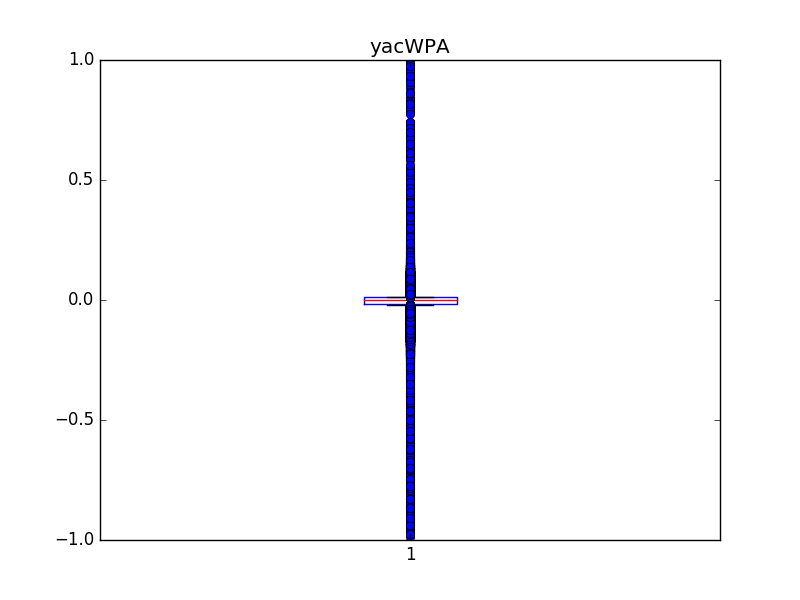
（9）yacWPA



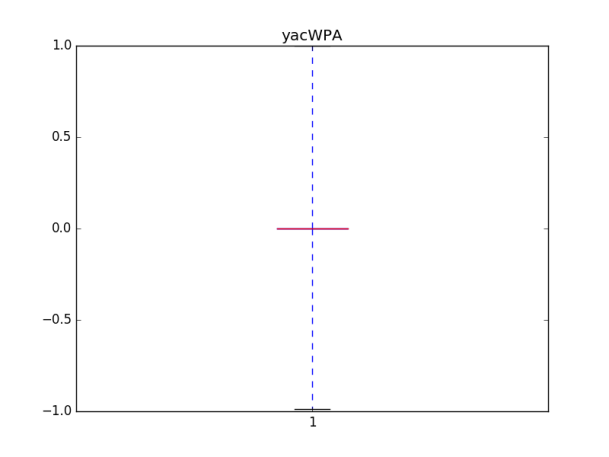
剔除缺失值 使用属性相关关系填充



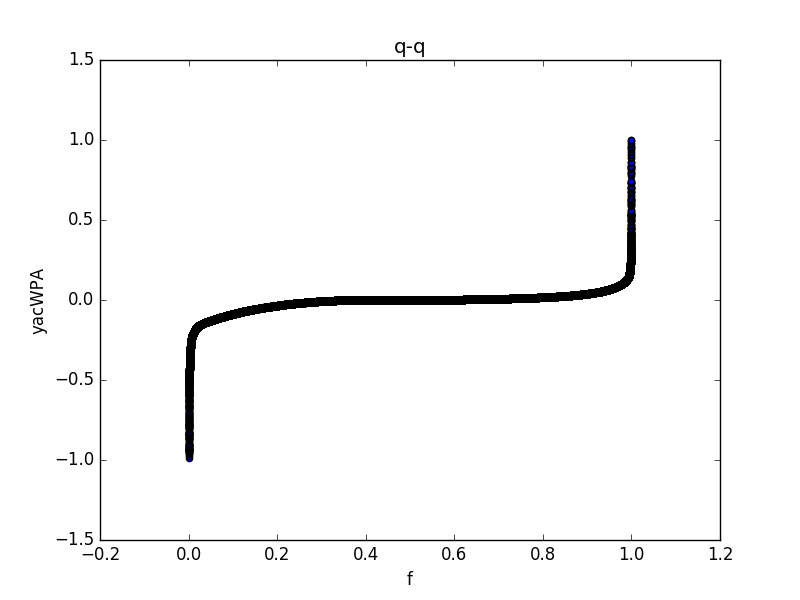
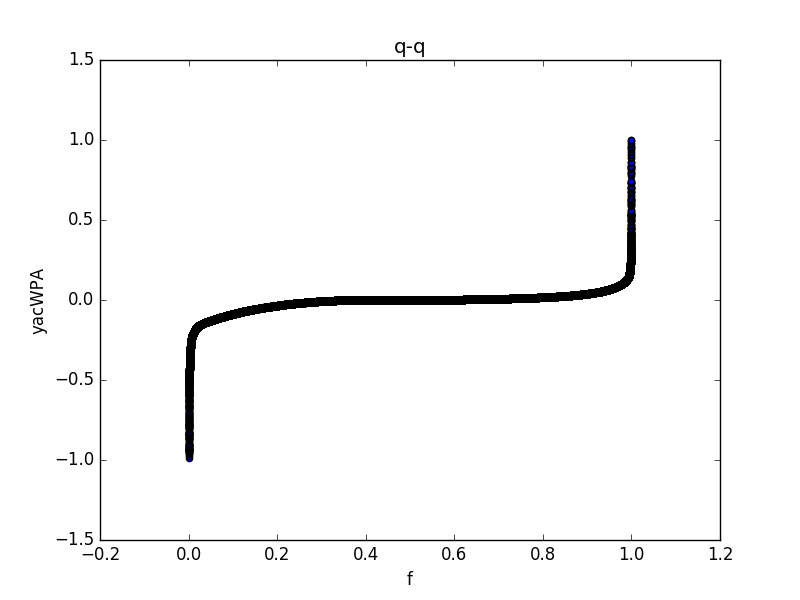
使用高频词填充



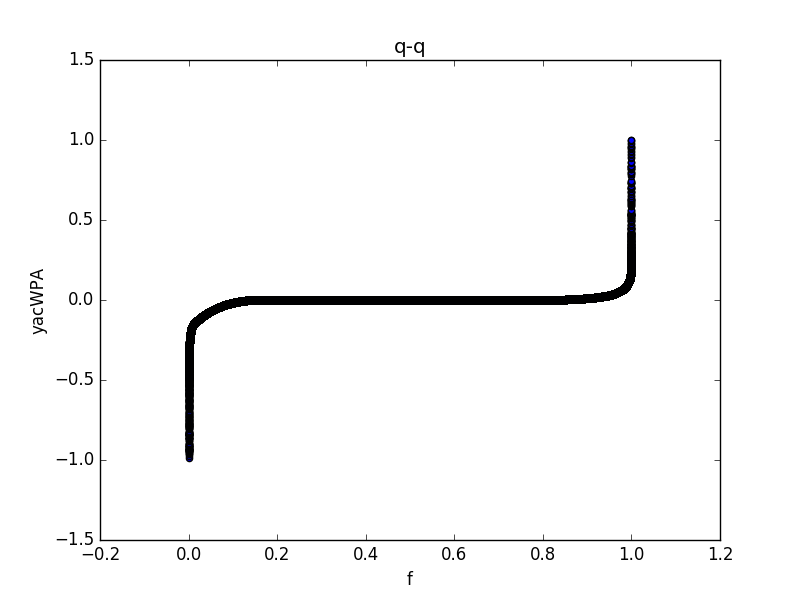
剔除缺失值 使用属性相关关系填充



使用高频词填充



剔除缺失值 使用属性相关关系填充



使用高频词填充

### 数据集2：NFL Play-by-Play 2009-2017

#### 数据摘要：

**1.标称属性：**Permit Number、Permit Type、Permit Type Definition、block、Lot、Street Number、Street Number Suffix、Street Name、Street Suffix、Unit、Unit Suffix、Description、Current Status、Structural Notification、Voluntary Soft-Story Retrofit、

Existing Use、Proposed Use、Existing Construction Type Description、Existing Construction Type、Proposed Construction Type、Supervisor District、Neighborhoods - Analysis Boundaries、Zipcode（邮政编码）、Record ID、Permit Creation Date、Current Status Date、Filed Date、Issued Date、First Construction Document Date、Permit Expiration Date

1. Permit Type

1:349 3:14663 2:950 5:91 4:2892 7:511 6:600 8:178844

1. Permit Type Definition

otc alterations permit:178844

new construction wood frame:950

sign - erect:2892

additions alterations or repairs:14663

grade or quarry or fill or excavate:91

demolitions:600

new construction:349

wall or painted sign:511

（3）Street Number Suffix

:196684 A:1501 C:56 B:291 E:28 D:11 P:3 F:24 I:7 H:11 K:11

J:9 L:10 N:2 G:12 0:1 R:10 V:228 ½:1

（4）Unit Suffix

:196939 -A:4 BLDG 3F:2 B20H:1 HMETER:1 BLDG 3E:4 PH1:1 M202:2

3/F:1 -B:6 TH2:2 TH1:5 C-2001:9 407:2 0:8 2ND FLOOR:1 FRONT:12

PH C:1 3RD FLOOR:1 BLDG 449:3 A104:2 B-A:5 A102:2 D:135 A100:1 H:23

L:5 P:13 BLDG 201:1 T:16 BASEMENT:1 C-4:1 PH:2 C-1:18 1ST FLOOR:3

REAR BLDG:23 COM'L:3 COMML-1:2 LOWER UNIT:4 FRNT RIGHT:4

STUDIO/TEC:1 TENTATIVE:2 PH2B:10 W122:1 REAR UNIT:7 Parcel B:1 Parcel C:1

M301:6 Parcel A:1 M1:3 REAR G/F:3 Parcel D:1 1503B:2 CML A:1

UPPR COMML:5 2/F:3 I:1 FRNT BLDG:9 PH1E:1 COMML-2:2 COMML-3:4

BLDG 229:2 C:226 G:36 K:10 PH1C:2 O:8 GPHA:2 MIDDLE:5 S:8

PH1A:2 W:43 COM'L 1/F:1 FRONT BLDG:23 M101:1 Garage:13 U:1 PH-4:1

RETAIL 1:5 C101:4 C102:11 RETAIL 2:1 UPPER UNIT:7 OFFICE:2

OFFICE 2/F:2 RESIDENCE:1 COMML 1:3 COMMERCIAL:10 RESIDENTIA:7

LOWER COML:2 REAR:11 BLDG 157:3 RES'L:1 RESDL/REAR:1 Commercial:8

1/F FRONT:2 1/F:16 BLDG E:1 BLDG D:1 BLDG C:1 RESIDL 3/F:3

BLDG A:1 BLD 7:2 CHILD DEV:1 EAST:2 BLDG B:1 FC1:1 B:191 BLDG 3:4

1/2:2 BLDG 1:12 F:78 J:19 PIER 2:1 PIER 3:1 PIER 1:1 R:14

COML:1 U-PH3:1 C1:2 UPPER:23 RESID'L:33 LEVEL ONE:1 PARCEL A:1

3/F FRONT:1 PH-1:8 UTILITY:1 FRNT ENTRA:2 LEVEL 4:1 LOWER:7 C100:4

FRONT UNIT:10 BLDG. 3:1 TOP:1 COMMON ARE:1 BLDG 2:1 1A-4:1 COM-1:1

C132:4 UPPER LEVE:1 COMML:43 BSM OFFICE:1 UP FRT BLD:1 A:346

E:92 HOA:85 LEAS OFFIC:6 M:8 COMM':1 LEASE OFF:2 Y:3 N:18 CU-3:2

CU-2:1 STORE:1 #R-2:1 R-1:6 THEATRE:4 WEST:8 COM-2:3 C-2:2 PARKING:2

D4:1 C103:3 D3:1

1. Current Status

suspend:193 complete:97077 issued:83559 plancheck:16 withdrawn:1754

reinstated:563 filed:12043 cancelled:1536 disapproved:2 appeal:2

expired:1370 approved:733 incomplete:2 revoked:50

1. Structural Notification

:191978 Y:6922

1. Voluntary Soft-Story Retrofit

:198865 Y:35

（7）Existing Use

:41114 office:24616 warehouse, furniture:99 greenhouse:27 workshop ommercial:189 recreation bldg:179 1 family dwelling:46766 swimming pool:5

child care:1 storage tanks:6 prkng garage/public:207

public assmbly other:196 nursery(floral):9 misc group residns.:113

automobile sales:95 ambulance service:3 school:674 day care home lt 7:9

storage shed:113 2 family dwelling:20987 tower:10 apartments:40798

prson'l svc tutor:5 library:33 sign:5 nursing home gt 6:13

day care home gt 12:53 sound studio:18 mortuary:15 tourist hotel/motel:1665

auto repairs:336 clinics-medic/dental:581 phone xchnge/equip:36

adult entertainment:9 parking lot:76 animal sale or care:59

day care home 7 - 12:4 nursing home non amb:10 manufacturing:510

muni carbarn:37 temple:2 orphanage:1 barber/beauty salon:212

lending institution:351 radio & tv stations:35 museum:146 wholesale sales:27

dry cleaners:63 moving & storage:30 massage parlor:78 vacant lot:1813

sewage plant:4 warehouse,no frnitur:870 church:467 roofing materials:1

accessory cottage:5 day care, non-res:41 car wash:11 hospital:38

dance hall:24 nursing home lte 6:5 health studios & gym:239 stadium:4

christmas tree lot:18 residential hotel:1020 building materials:10

social care facility:43 sfpd or sffd station:97 club:124

workshop residential:5 meat/produce marts:5 day care center:27 paint store:5

printing plant:69 dairies/dairy equip.:1 garment shops:40 nite club:79

muni driver restroom:3 laundry/laundromat:99 fence/retaining wall:9

convalescent home:17 antenna:110 chemical processing:9 r-3(dwg) nursing:4

filling/service stn:218 bath house:2 retail sales:6910 power plant:68

artist live/work:448 amusement center:7 theater:219

prkng garage/private:194 jail:6 food/beverage hndlng:4886

（8）Proposed Use

:42439 office:23962 warehouse, furniture:65 greenhouse:72

workshop commercial:219 recreation bldg:207 1 family dwelling:46346

swimming pool:6 child care:25 storage tanks:5 prkng garage/public:175

public assmbly other:228 nursery(floral):25 misc group residns.:119

automobile sales:74 ambulance service:4 not applicable:1 school:748

day care home lt 7:11 storage shed:79 2 family dwelling:22061 tower:11

apartments:43032 prson'l svc tutor:15 library:33 sign:6

nursing home gt 6:14 day care home gt 12:83 bath house:5 mortuary:12

tourist hotel/motel:1601 auto repairs:239 clinics-medic/dental:678

phone xchnge/equip:38 adult entertainment:5 parking lot:67

animal sale or care:60 day care home 7 - 12:11 nursing home non amb:11

manufacturing:436 muni carbarn:35 temple:2 orphanage:1

barber/beauty salon:258 lending institution:239 radio & tv stations:36

hospital:32 wholesale sales:40 dry cleaners:36 moving & storage:26

massage parlor:104 vacant lot:385 sewage plant:6 warehouse,no frnitur:674

church:434 roofing materials:1 accessory cottage:5 day care, non-res:59

car wash:5 museum:139 dance hall:22 nursing home lte 6:4

health studios & gym:332 stadium:3 christmas tree lot:18

residential hotel:1002 building materials:6 social care facility:47

sfpd or sffd station:98 club:125 workshop residential:10

meat/produce marts:2 day care center:40 paint store:6 printing plant:34

dairies/dairy equip.:1 garment shops:24 nite club:65

muni driver restroom:5 laundry/laundromat:56 fence/retaining wall:22

convalescent home:16 antenna:115 chemical processing:9 r-3(dwg) nursing:4

filling/service stn:99 sound studio:28 retail sales:5079 power plant:82

artist live/work:443 amusement center:8 theater:189

prkng garage/private:102 jail:6 food/beverage hndlng:5053

1. Existing Construction Type

:43366 1:28072 3:9663 2:4068 5:113350 4:381

1. Proposed Construction Type

:43162 1:27841 3:9360 2:3778 5:114382 4:377

1. Supervisor District

11:6940 10:12152 :1717 1:13038 3:28649 2:25483 5:19045 4:9592 7:14365

6:24797 9:16362 8:26760

1. Neighborhoods - Analysis Boundaries

:1725 Inner Richmond:4458 Japantown:700 West of Twin Peaks:8739

Glen Park:2637 Western Addition:3867 Outer Richmond:7854 Inner Sunset:5776

Financial District/South Beach:21816 Excelsior:3072 Presidio Heights:4084

Bernal Heights:6067 Golden Gate Park:64 South of Market:7572 Portola:1433

Potrero Hill:4293 Treasure Island:81 Pacific Heights:8508

Hayes Valley:5489 Oceanview/Merced/Ingleside:2654 Visitacion Valley:900

Marina:8244 Sunset/Parkside:10207 McLaren Park:46 Nob Hill:6009

Lakeshore:1308 Mission Bay:2287 Chinatown:3765 Noe Valley:7844

North Beach:4054 Castro/Upper Market:8527 Haight Ashbury:5798

Lincoln Park:49 Bayview Hunters Point:5669 Seacliff:992 Lone Mountain/USF:3358

Presidio:51 Mission:14681 Tenderloin:4783 Russian Hill:5495

Outer Mission:2242 Twin Peaks:1702

1. Zipcode

:1716 94130:81 94131:7664 94132:3507 94133:7424 94134:2983 94158:1058

94118:9812 94112:7897 94110:17837 94111:5385 94116:6421 94117:11780

94114:13404 94115:10095 94127:4993 94124:5265 94123:9515 94122:8886

94121:7773 94129:23 94109:11348 94108:5320 94103:10986 94102:7164

94105:8628 94104:4229 94107:7706

1. **数值属性：**Number of Existing Stories、Number of Proposed Stories、estimated cost、Revised Cost、Existing Units、Proposed Units、Plansets

Number of Existing Stories:

最大值:78.00 最小值:0.00 均值:0.00 中位数:3.00 四分位数\_1:2.00 四分位数\_3:4.00 缺失值:42784

Number of Proposed Stories:

最大值:78.00 最小值:0.00 均值:0.00 中位数:3.00 四分位数\_1:2.00 四分位数\_3:4.00 缺失值:42868

Estimated Cost:

最大值:537958646.00 最小值:1.00 均值:1.00 中位数:11000.00 四分位数\_1:3300.00 四分位数\_3:35000.00 缺失值:38066

Revised Cost:

最大值:780500000.00 最小值:0.00 均值:0.00 中位数:7000.00 四分位数\_1:1.00 四分位数\_3:28710.00 缺失值:6066

Existing Units:

最大值:1907.00 最小值:0.00 均值:0.00 中位数:1.00 四分位数\_1:1.00 四分位数\_3:4.00 缺失值:51538

Proposed Units:

最大值:1911.00 最小值:0.00 均值:0.00 中位数:2.00 四分位数\_1:1.00 四分位数\_3:4.00 缺失值:50911

Plansets:

最大值:9000.00 最小值:0.00 均值:0.00 中位数:2.00 四分位数\_1:0.00 四分位数\_3:2.00 缺失值:37309

#### 数据可视化：

数值属性之间的相似性为：

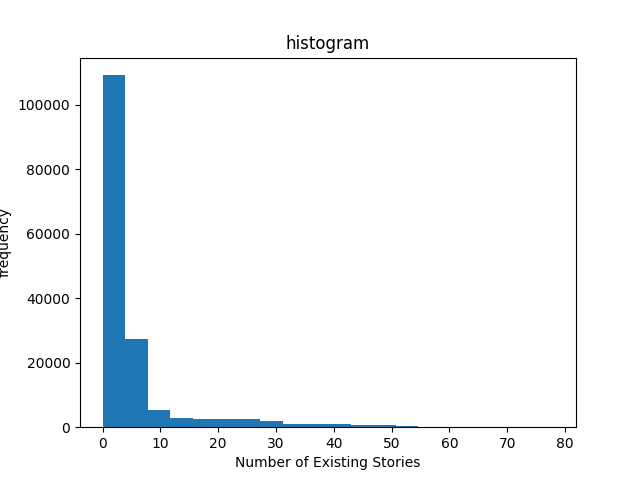
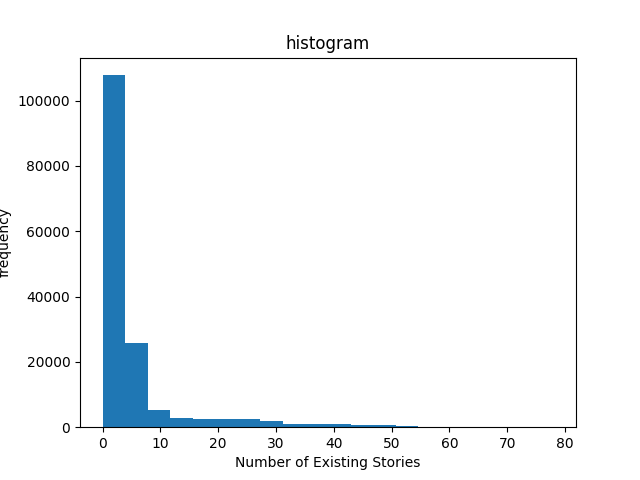
Number of Existing Stories相似于Number of Proposed Stories

estimated cost相似于Revised Cost

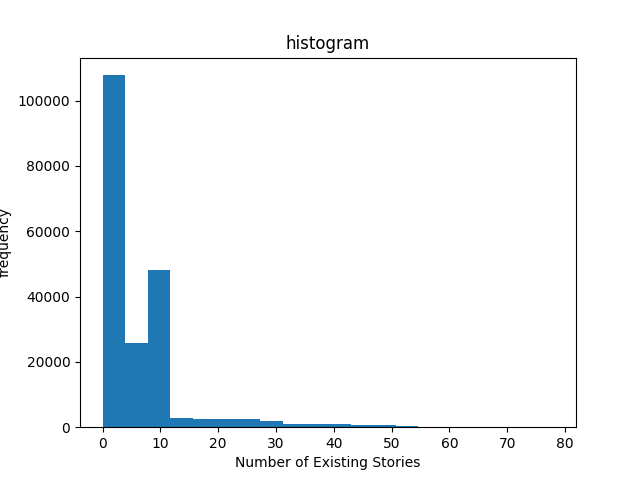
Existing Units相似于Proposed Units

可以从下面的图片中看到，尽管使用了不同的填充方法填充缺失值，但因为数据数量很大，而缺失值只占一小部分，所以使用不同的填充方法画出来的结果是大体相似的

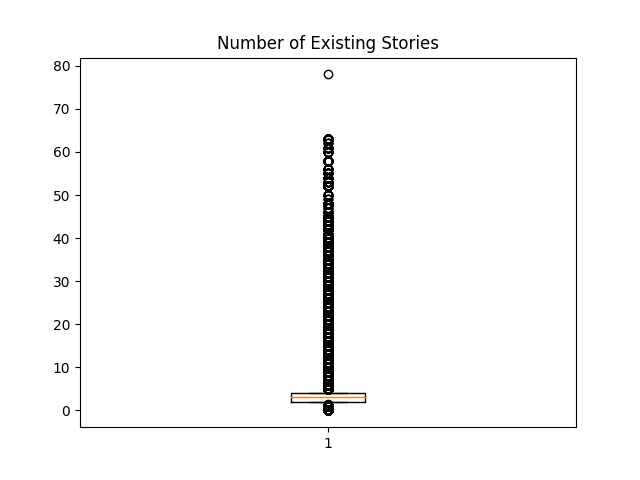
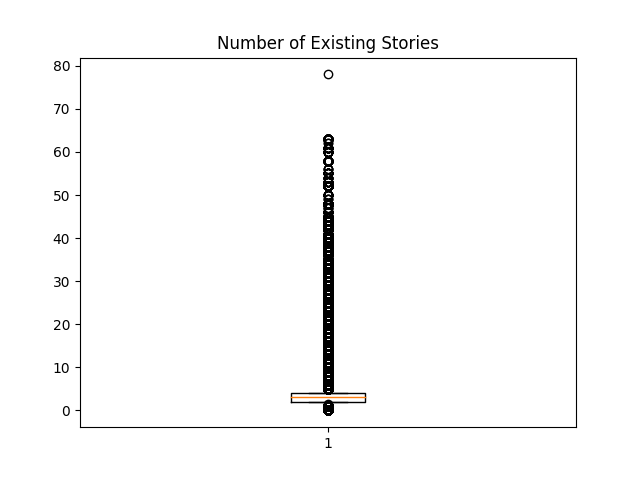
（1）Number of Existing Stories



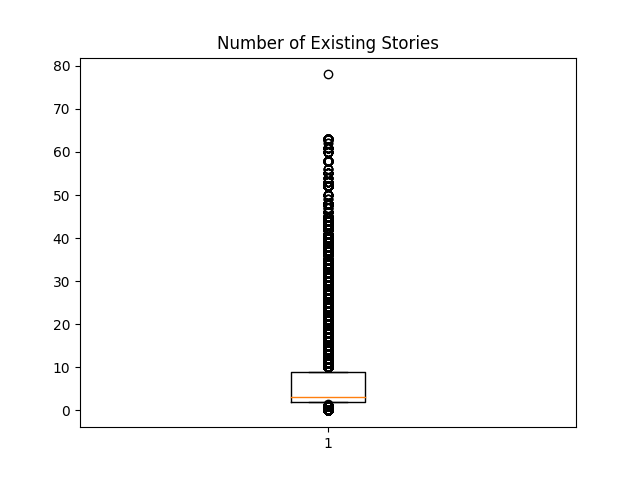
剔除缺失值 属性间相似关系填充



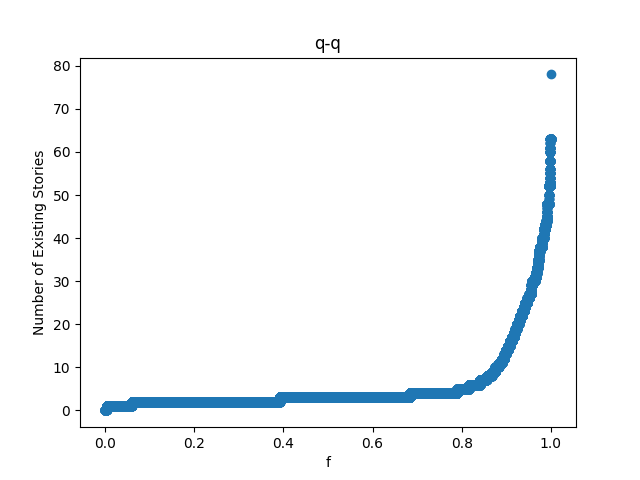
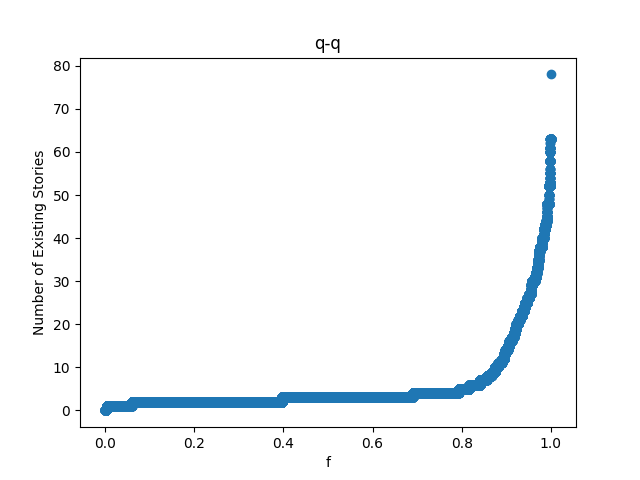
用高频词填充



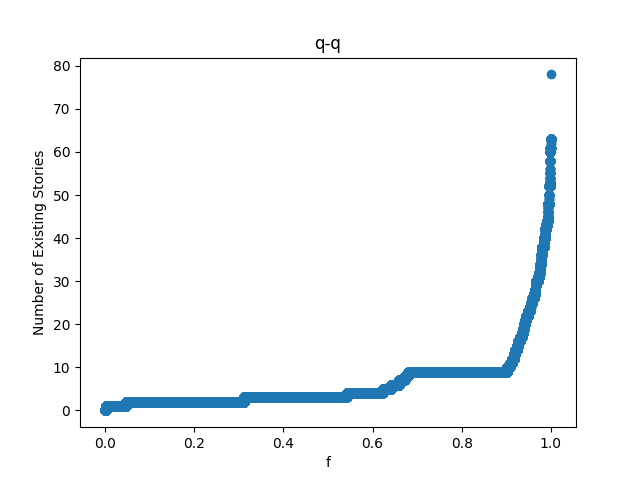
剔除缺失值 属性间相似关系填充



用高频词填充

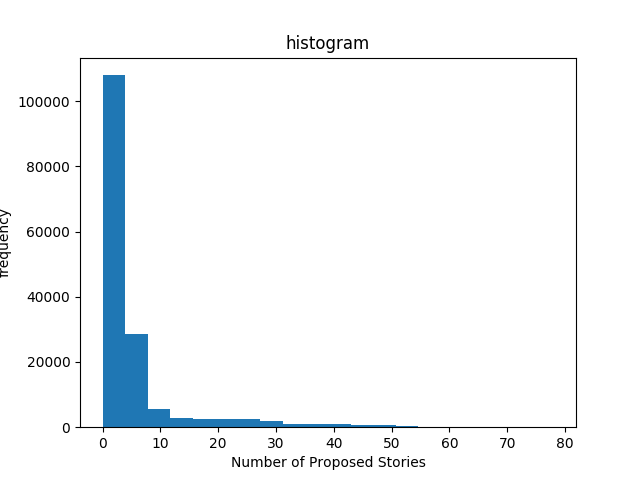
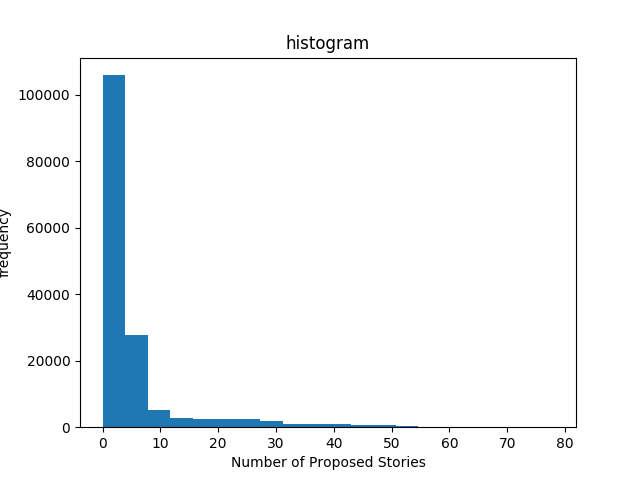


剔除缺失值 属性间相似关系填充

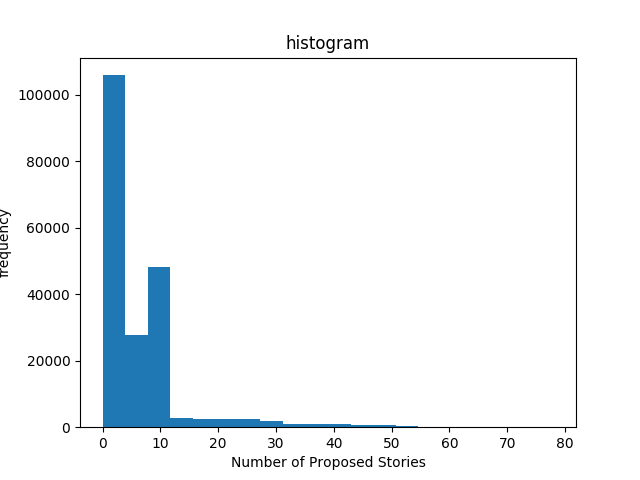


用高频词填充

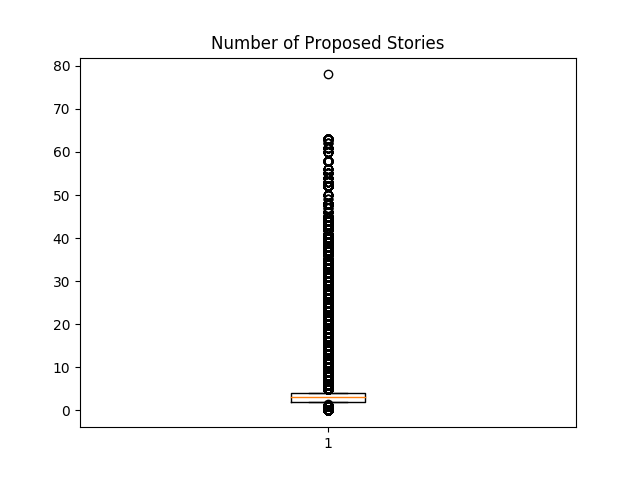
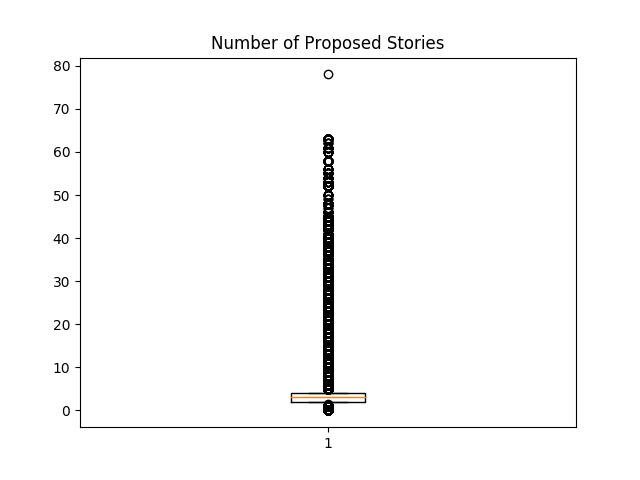
（2）Number of Proposed Stories



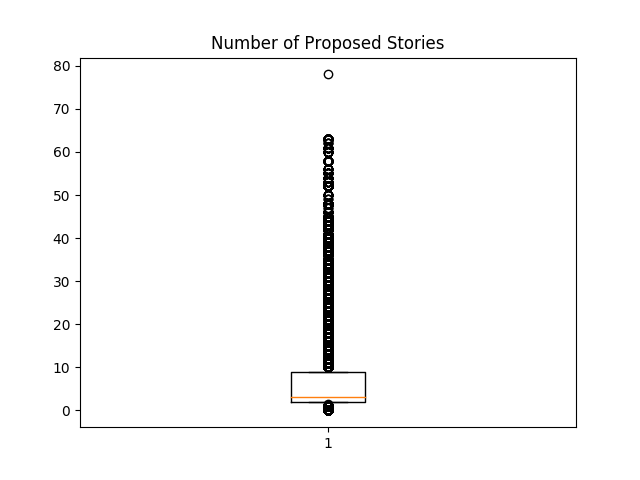
剔除缺失值 属性的相似关系填充

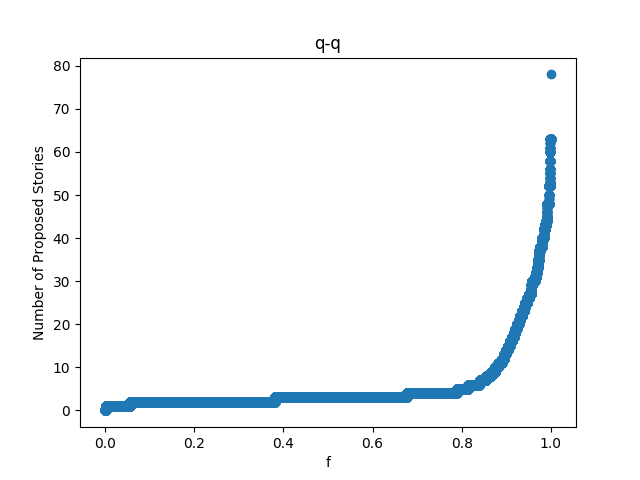
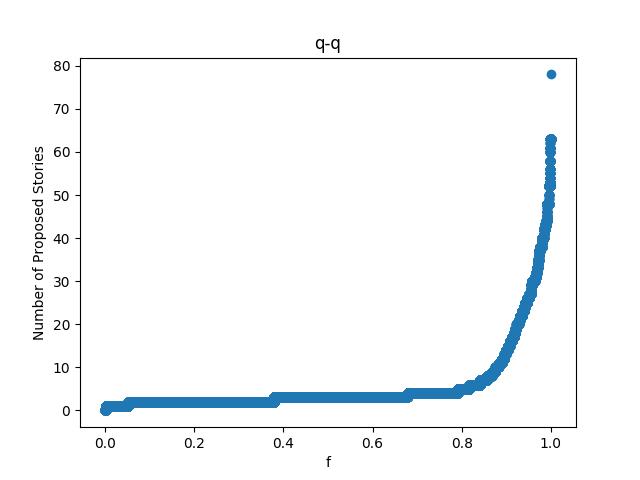


使用高频词填充

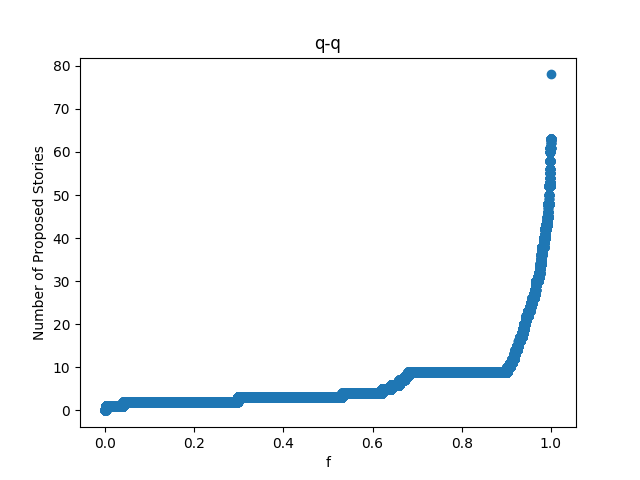


剔除缺失值 属性的相似关系填充

使用高频词填充

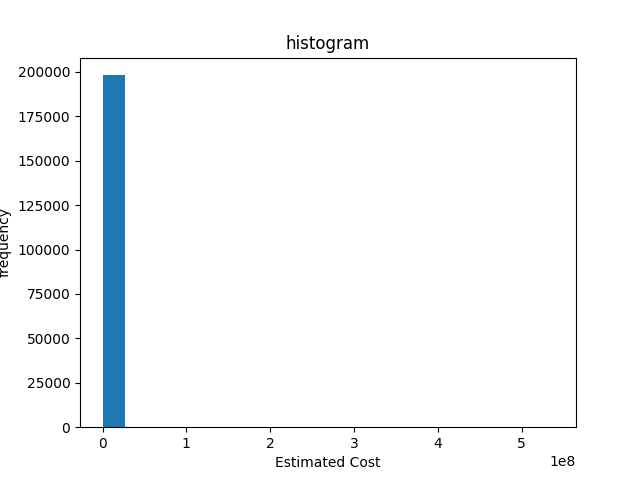
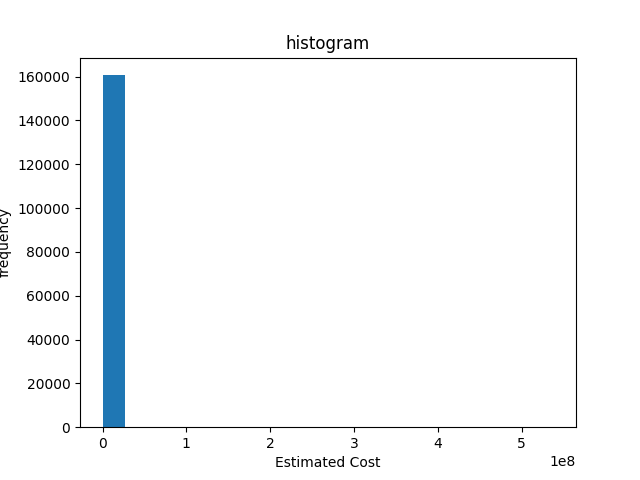


剔除缺失值 属性的相似关系填充

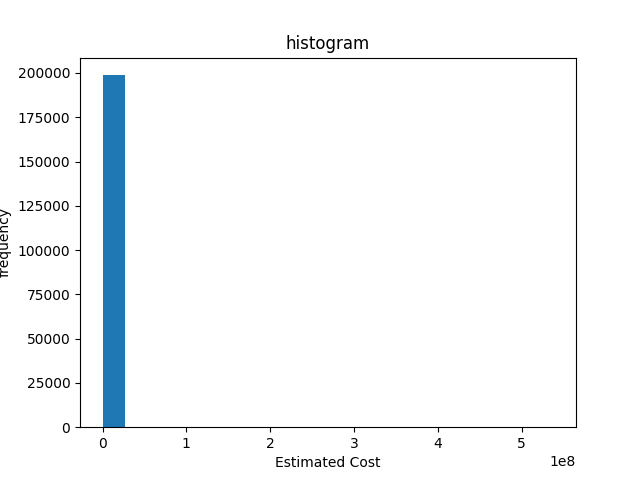


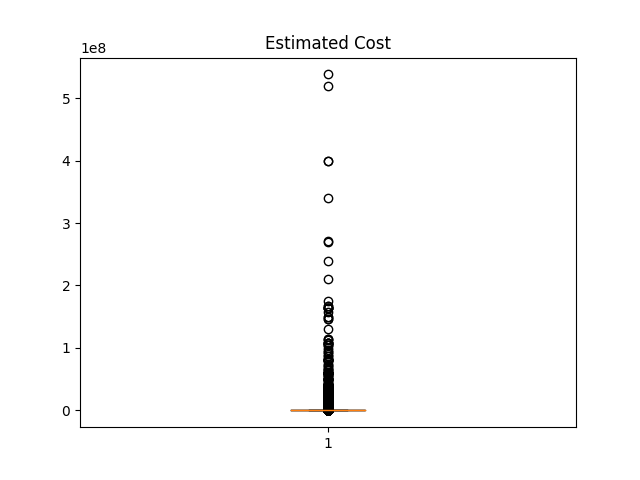
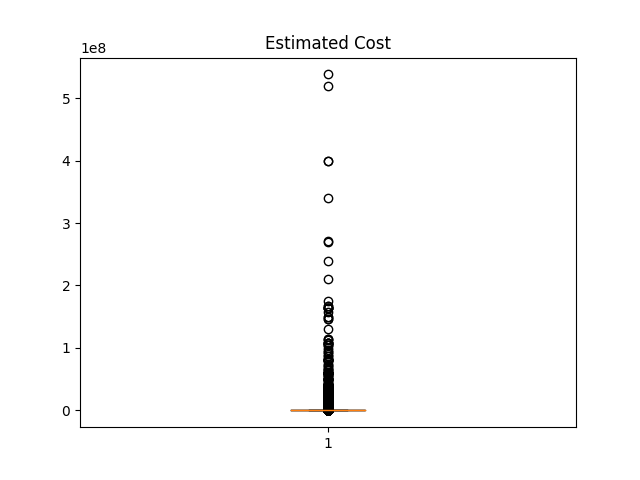
使用高频词填充

1. estimated cost

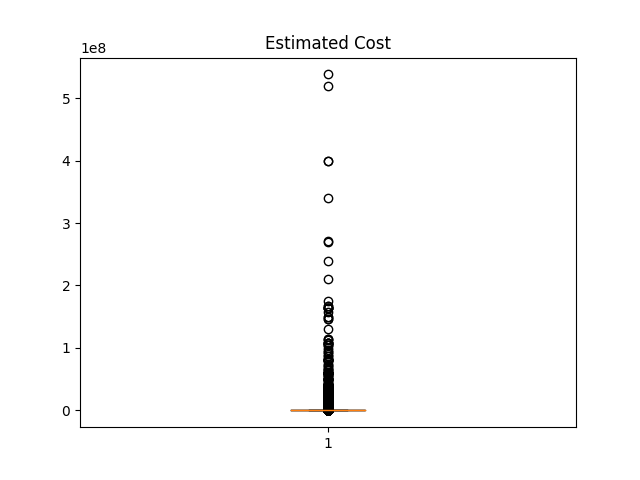


使用缺失值填充 属性的相关关系

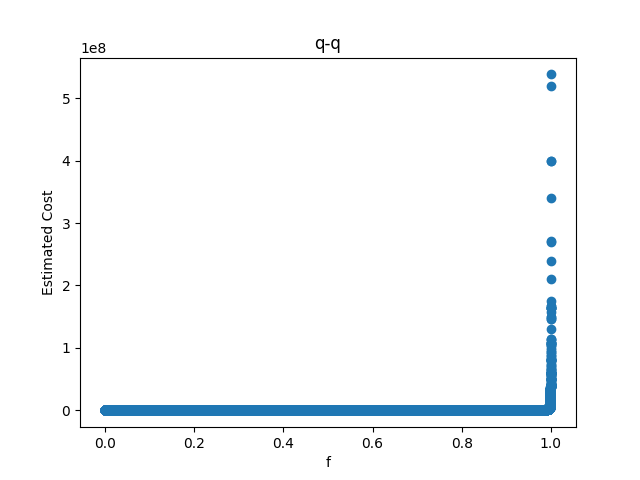
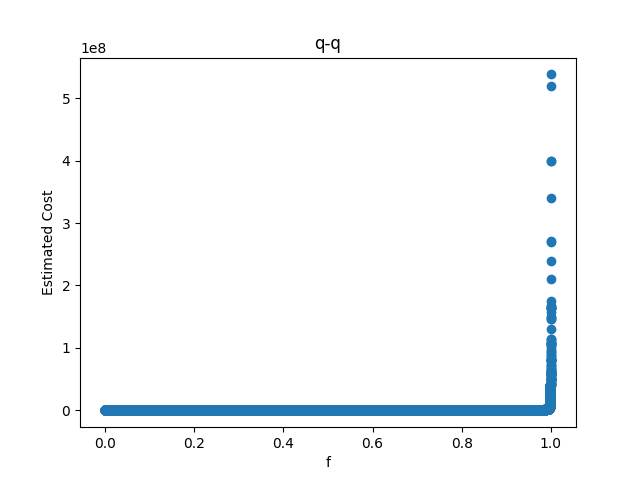
使用高频词填充



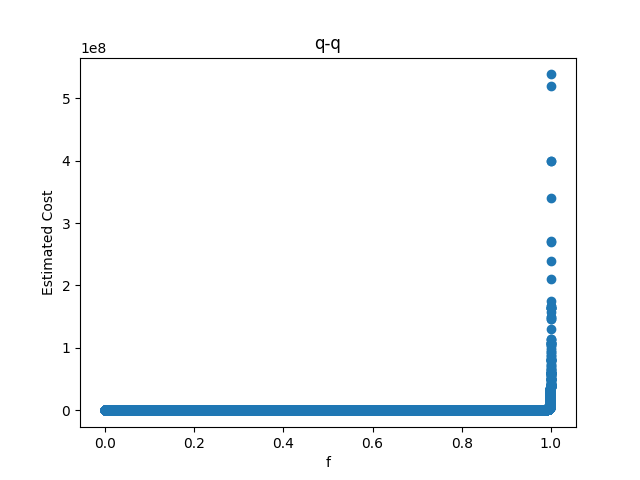
使用缺失值填充 属性的相关关系



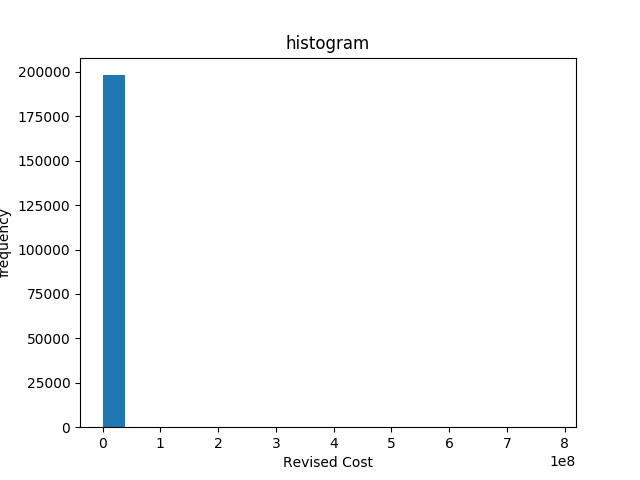
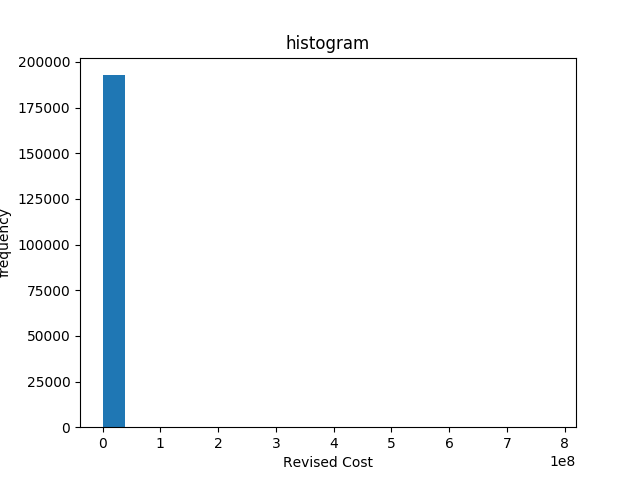
使用高频词填充



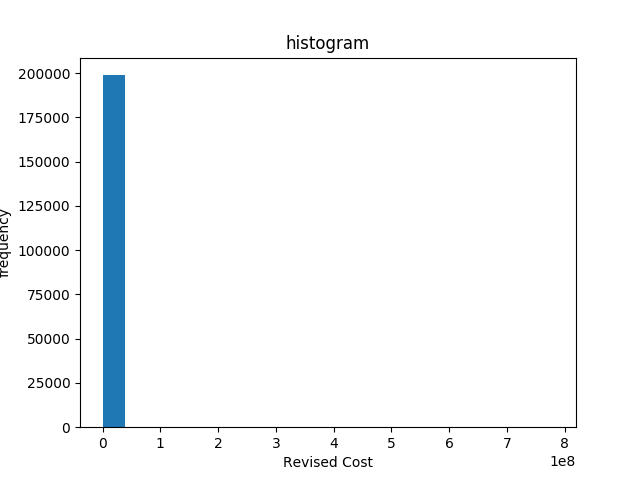
使用缺失值填充 属性的相关关系

使用高频词填充

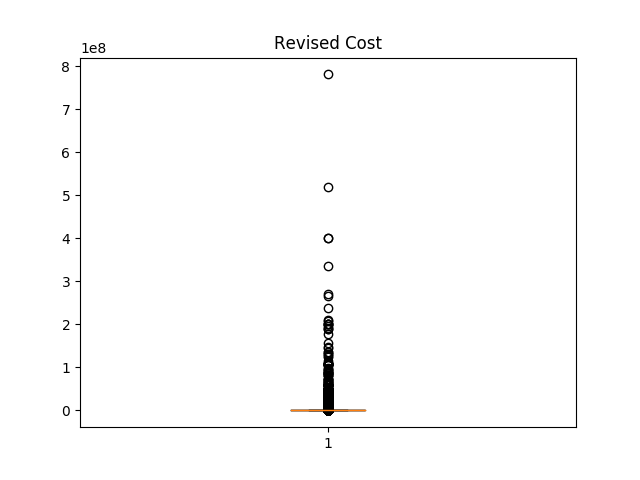
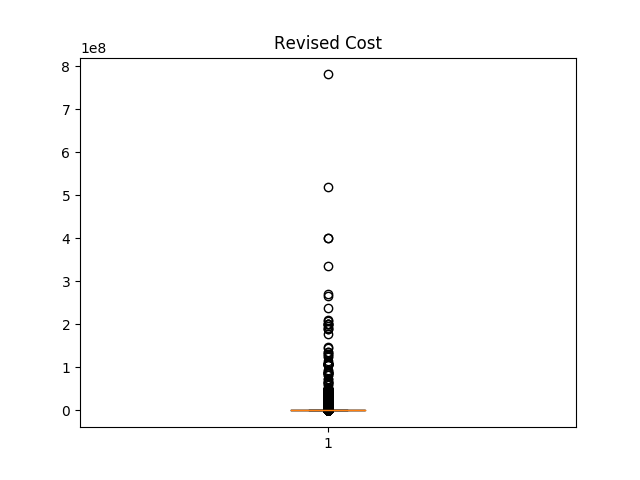
（4）Revised Cost



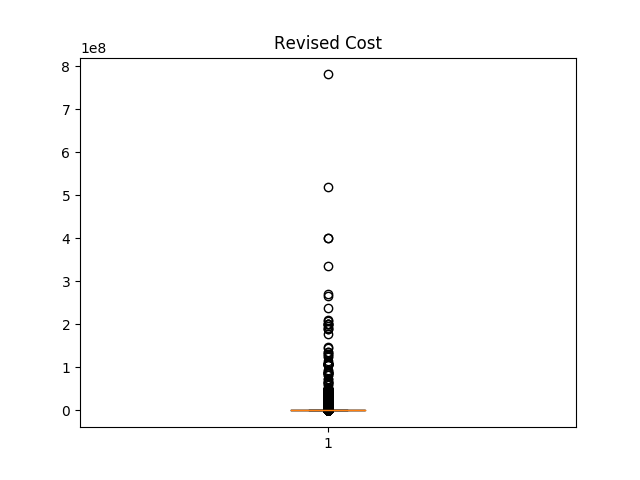
剔除缺失值 属性相似关系填充



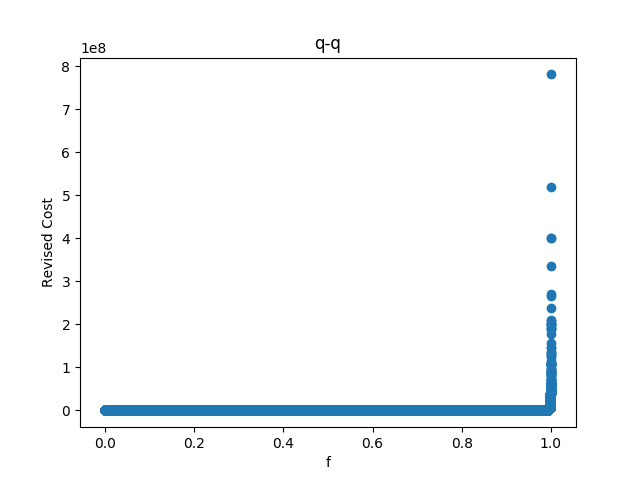
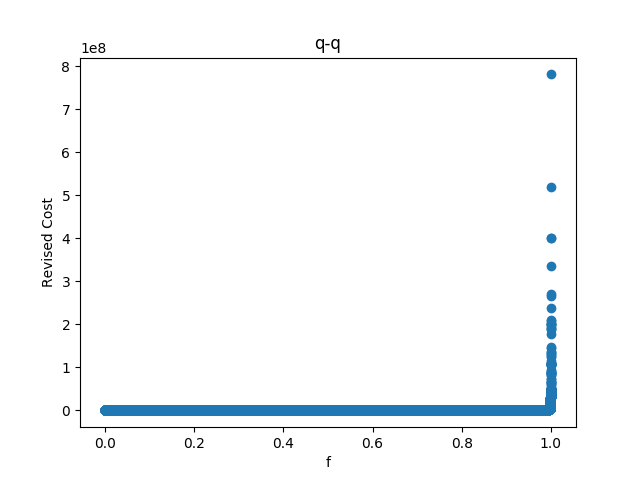
高频词填充



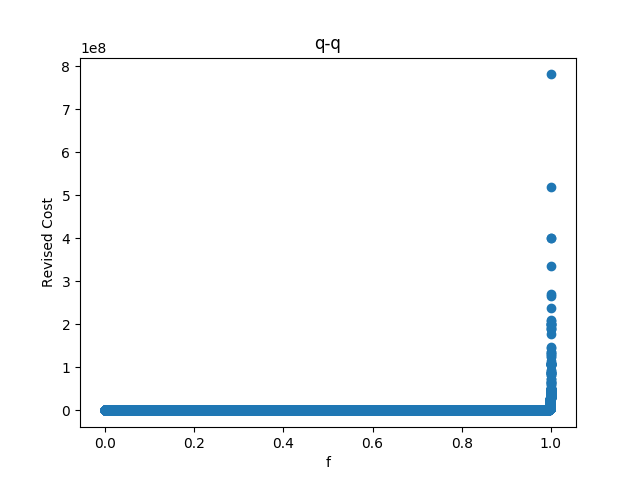
剔除缺失值 属性相似关系填充



高频词填充

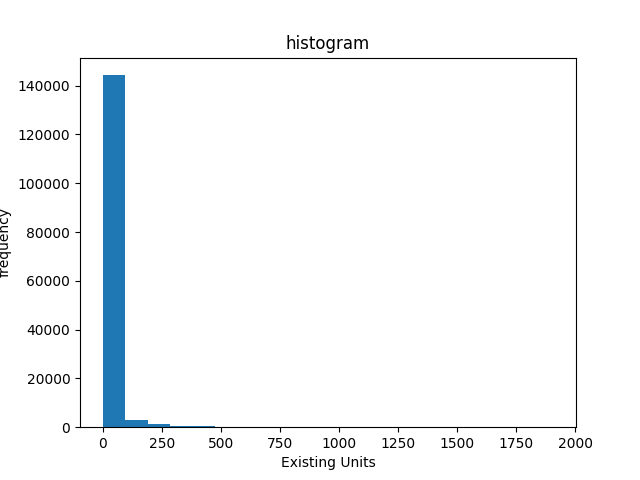
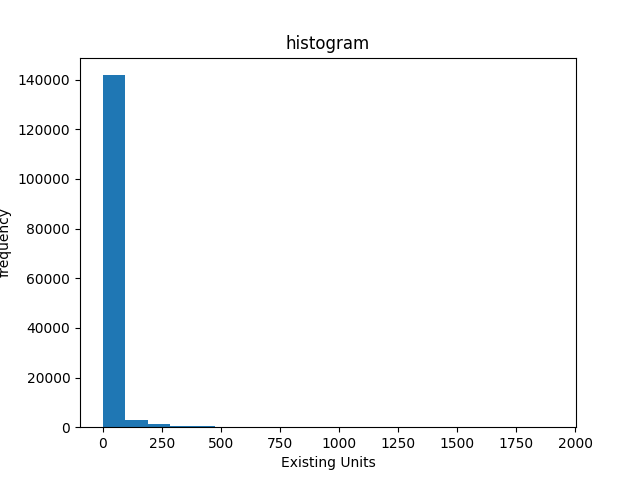


剔除缺失值 属性相似关系填充

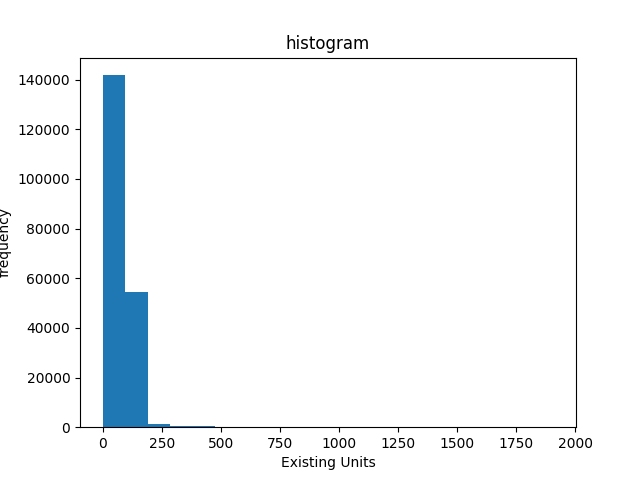


高频词填充

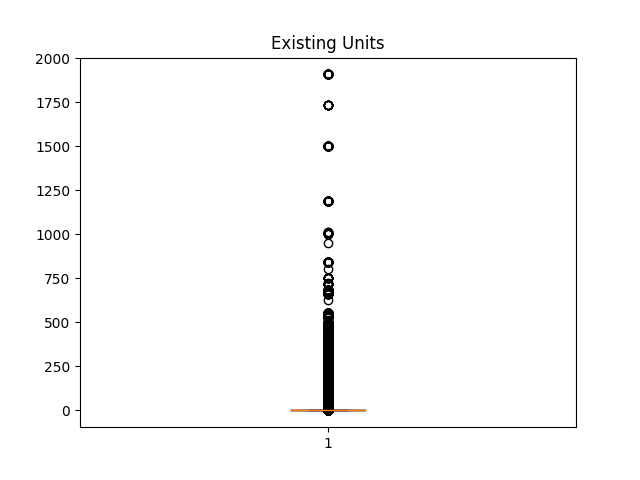
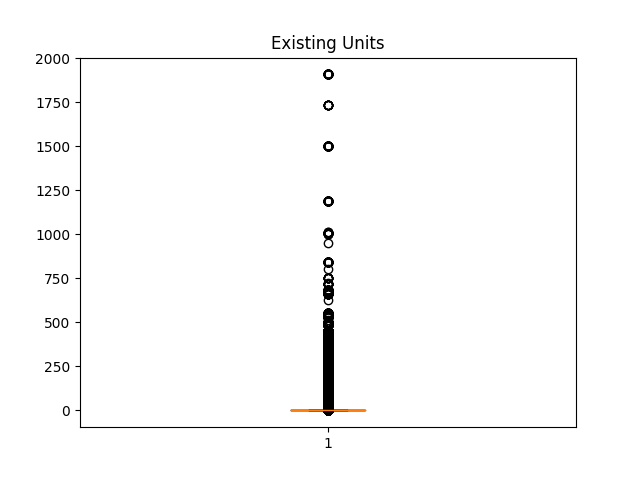
（5）Existing Units



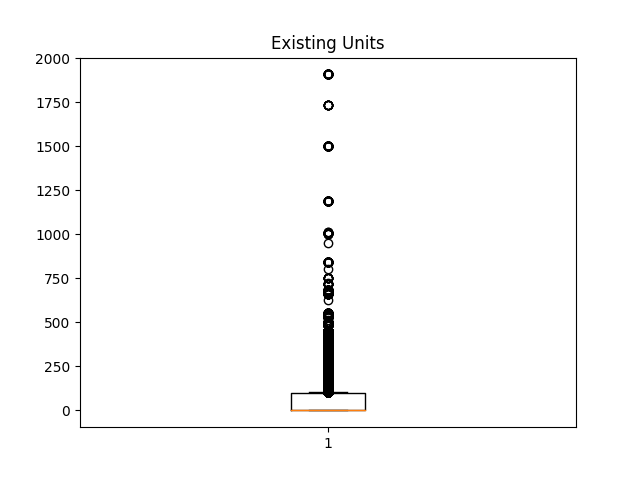
剔除缺失值 属性间的相似性



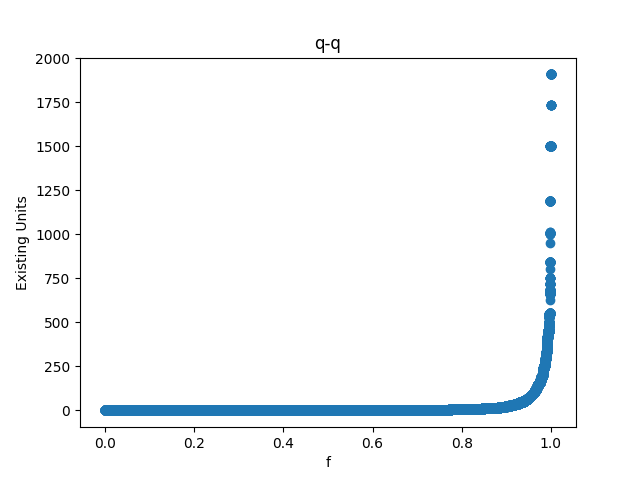
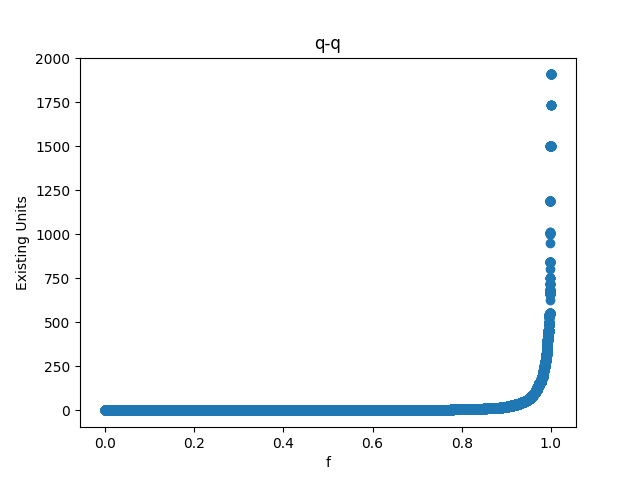
使用高频词填充



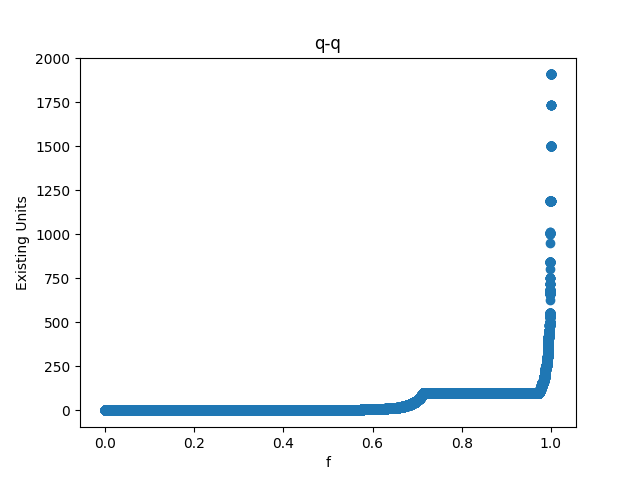
剔除缺失值 属性间的相似性



使用高频词填充

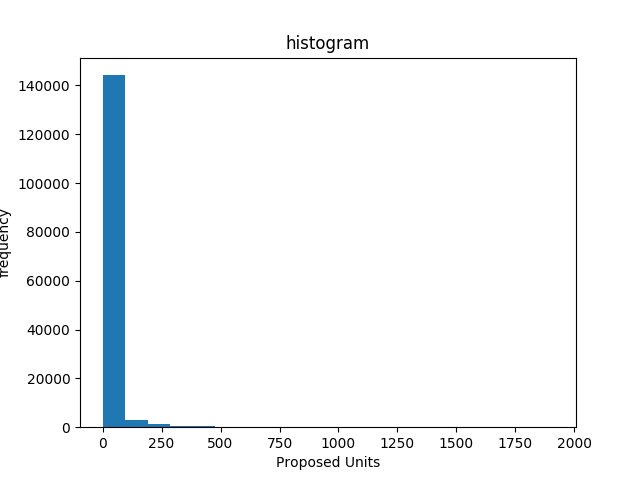
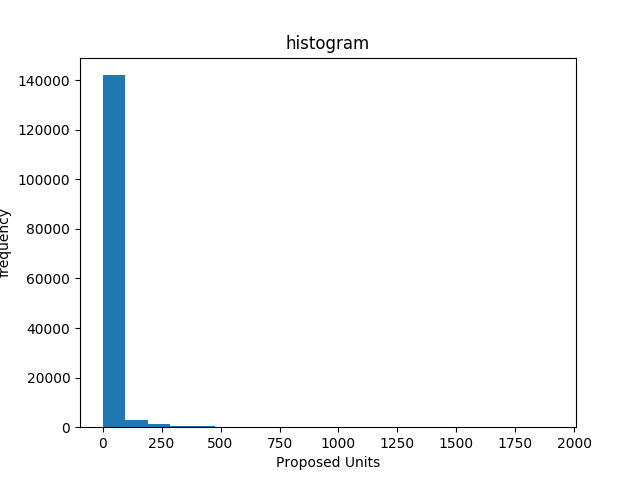


剔除缺失值 属性间的相似性

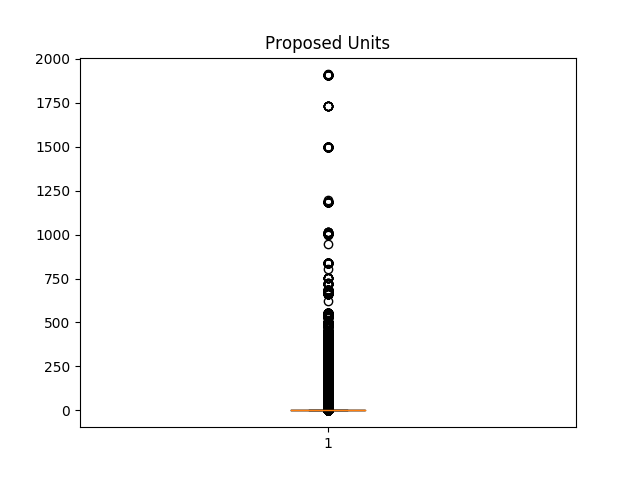
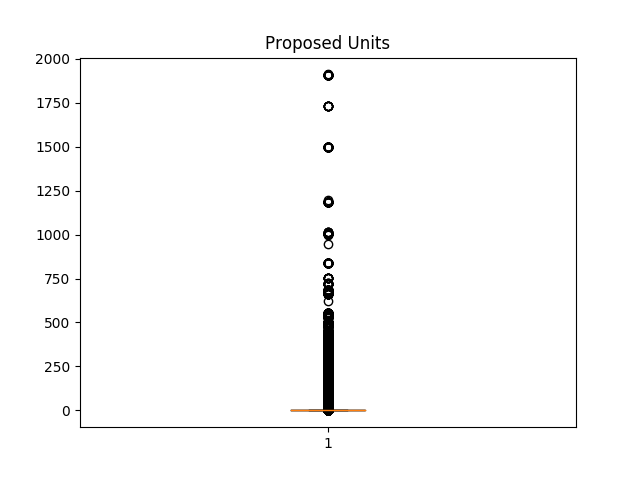


使用高频词填充

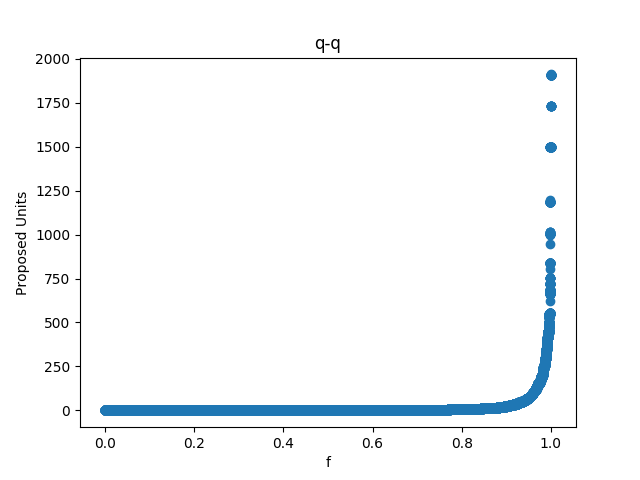
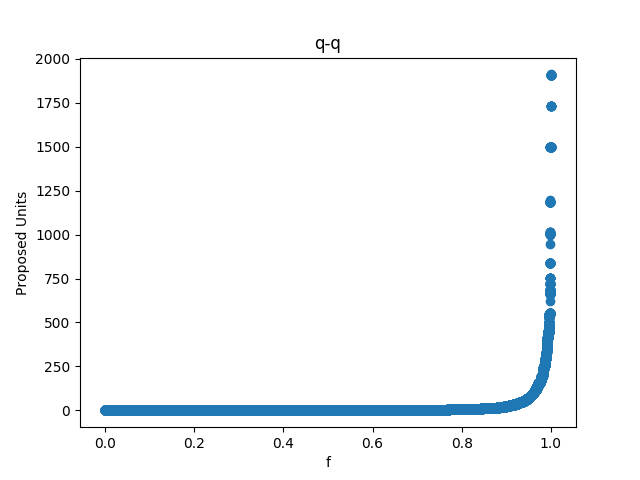
（6）Proposed Units



剔除缺失值 使用相似属性填充



剔除缺失值 使用相似属性填充



剔除缺失值 使用相似属性填充

#### 代码实现：

两个数据集的代码是相似的，下面只展示一个数据集上的代码。

统计标称属性代码：

from \_\_future\_\_ import print\_function

import csv

import os

filename='./biaocheng/'

with open('Building\_Permits.csv') as cs:

csv\_reader=list(csv.reader(cs))

labels=csv\_reader[0]

lenth=len(labels)

l=len(csv\_reader)

for i in range(lenth):

filename1=filename+labels[i]+'.txt'

result={}

for row in csv\_reader[1:l]:

if not result.has\_key(row[i]):

result[row[i]]=0

result[row[i]]+=1

keys=result.keys()

with open(filename1,'w') as f:

for key in keys:

f.write("%s:%d\n"%(key,result[key]))

f.close()

cs.close()

统计数值属性最大值、最小值等属性的代码：

#coding=utf-8

from \_\_future\_\_ import print\_function

import csv

import os

#比率标度属性

label\_shuzhi=[

'Number of Existing Stories',

'Number of Proposed Stories',

'Estimated Cost',

'Revised Cost',

'Existing Units',

'Proposed Units',

'Plansets'

]

filename='shuzhi.txt'

with open('Building\_Permits.csv') as cs:

with open(filename,'w') as f:

csv\_reader=list(csv.reader(cs))

labels=csv\_reader[0]

lenth=len(labels)

for i in range(lenth):

#处理数值属性

if labels[i] in label\_shuzhi:

data=[]

queshi\_num=0

for row in csv\_reader:

if row[i]==labels[i]:

continue

elif row[i].strip()=='':

queshi\_num+=1

else:

data.append(float(row[i]))

else:

continue

f.write("%s:\n"%labels[i])

if len(data)>0:

len1=len(data)

sorted\_data=sorted(data)

max\_mum=max(data)#最大值

min\_mum=min(data)#最小值

mid\_mum=sum(data)/float(len1)#均值

mid\_num=sorted\_data[len1/2]#中位数

quater\_1=sorted\_data[len1/4]#四分为数

quater\_3=sorted\_data[len1-len1/4]#四分位数

f.write("最大值:%.2f 最小值:%.2f 均值:%.2f 中位数:%.2f 四分位数\_1:%.2f 四分位数\_3:%.2f 缺失值:%d\n"%(max\_mum,min\_mum,min\_mum,mid\_num,quater\_1,quater\_3,queshi\_num))

f.close()

cs.close()

剔除缺失值画直方图代码：

#coding=utf-8

from \_\_future\_\_ import print\_function

import csv

import os

import matplotlib.pyplot as plt

#比率标度属性

label\_shuzhi=[

'Number of Existing Stories',

'Number of Proposed Stories',

'Estimated Cost',

'Revised Cost',

'Existing Units',

'Proposed Units',

'Plansets'

]

filename='./zhifangtu/'

with open('Building\_Permits.csv') as cs:

csv\_reader=list(csv.reader(cs))

labels=csv\_reader[0]

length=len(labels)

leng=len(csv\_reader)

for i in range(length):

#处理数值属性

if labels[i] not in label\_shuzhi:

continue

else:

data=[]

queshi\_num=0

for row in csv\_reader[1:leng]:

if row[i].strip()=='':

queshi\_num+=1

else:

data.append(float(row[i]))

fig=plt.figure()

ax=fig.add\_subplot(111)

ax.hist(data,bins=20)

plt.xlabel(labels[i])

plt.ylabel('frequency')

plt.title('histogram')

plt.savefig(filename+labels[i]+'.png')

cs.close()

剔除缺失值画盒图代码：

#coding=utf-8

from \_\_future\_\_ import print\_function

import csv

import os

import matplotlib.pyplot as plt

#比率标度属性

label\_shuzhi=[

'Number of Existing Stories',

'Number of Proposed Stories',

'Estimated Cost',

'Revised Cost',

'Existing Units',

'Proposed Units',

'Plansets'

]

filename='./hetu/'

with open('Building\_Permits.csv') as cs:

csv\_reader=list(csv.reader(cs))

labels=csv\_reader[0]

length=len(labels)

leng=len(csv\_reader)

for i in range(length):

#处理数值属性

if labels[i] not in label\_shuzhi:

continue

else:

data=[]

queshi\_num=0

for row in csv\_reader[1:leng]:

if row[i].strip()=='':

queshi\_num+=1

else:

data.append(float(row[i]))

fig=plt.figure()

ax=fig.add\_subplot(111)

ax.boxplot(data,sym='o',whis=0.05)

plt.title(labels[i])

plt.savefig(filename+labels[i]+'.png')

cs.close()

剔除缺失值画分位数图代码：

#coding=utf-8

from \_\_future\_\_ import print\_function

import csv

import os

import matplotlib.pyplot as plt

#比率标度属性

label\_shuzhi=[

'Number of Existing Stories',

'Number of Proposed Stories',

'Estimated Cost',

'Revised Cost',

'Existing Units',

'Proposed Units',

'Plansets'

]

filename='./qq/'

with open('Building\_Permits.csv') as cs:

csv\_reader=list(csv.reader(cs))

labels=csv\_reader[0]

length=len(labels)

leng=len(csv\_reader)

for i in range(length):

#处理数值属性

if labels[i] not in label\_shuzhi:

continue

else:

x=[]

y=[]

queshi\_num=0

for row in csv\_reader[1:leng]:

if row[i].strip()=='':

queshi\_num+=1

else:

y.append(float(row[i]))

y=sorted(y)

l=len(y)

for j in range(l):

x.append((j+1-0.5)/float(l))

fig=plt.figure()

ax=fig.add\_subplot(111)

ax.scatter(x,y)

plt.xlabel("f")

plt.ylabel(labels[i])

plt.title('q-q')

plt.savefig(filename+labels[i]+'.png')

cs.close()