Big Match - Large Demo

Rachael Caelie (Rocky) Aikens

10/30/2018

This rmarkdown is meant for running a larger demo of the big_match functionality.

Import Data

This sample cohort data was provided by Justin Lee at the Quantitative Sciences Unit. It contains ~900,000 observations of 112 variables.

```
dat <- read_sas("../sample_data/justincohort_june2017.sas7bdat")

# dimensions: ~900,000 x 112
dim(dat)
```

```
## [1] 893498    112
```

```
# filter and add treatment column
dat <- filter(dat, totalct > 1 & arteryCt < 3) %>%
  mutate(treat = ifelse(arteryCt > 1, 1, 0))

# dimensions: ~900,000 x 112
dim(dat)
```

```
## [1] 833657    113
```

Manual Statify

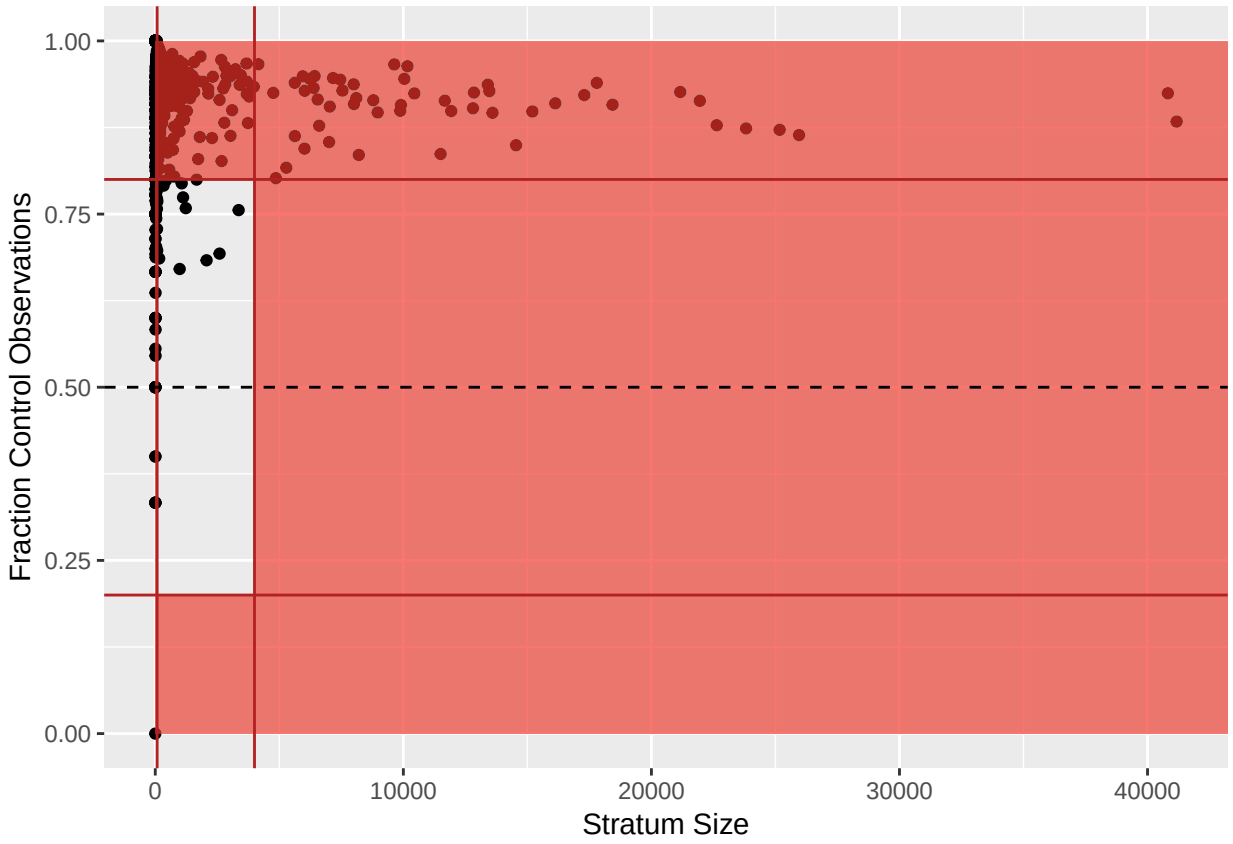
```
m.strat <- manual_stratify(dat, treat = "treat",
                           covariates = c("Male", "race", "hosp_state"))

summary(m.strat)
```

```
## $call
## manual_stratify(data = dat, treat = "treat", covariates = c("Male",
##     "race", "hosp_state"))
##
## $issue_table
## # A tibble: 704 x 6
##    Stratum Treat Control Total Control_Proporti~ Potential_Issues
##      <dbl> <dbl>   <dbl> <int>             <dbl> <chr>
##  1       1     0       3     3             1     Too few samples; Not eno~
##  2       2     0      13    13             1     Too few samples; Not eno~
##  3       3     0       6     6             1     Too few samples; Not eno~
##  4       4     1      19    20             0.95  Too few samples; Not eno~
##  5       5     0       5     5             1     Too few samples; Not eno~
##  6       6    64    1071  1135             0.944 Not enough treated sampl~
##  7       7     1      10    11             0.909 Too few samples; Not eno~
##  8       8     1      14    15             0.933 Too few samples; Not eno~
##  9       9     0       6     6             1     Too few samples; Not eno~
## 10      10     1      35    36             0.972 Too few samples; Not eno~
## # ... with 694 more rows
##
## $sum_before
##                          Treat_Mean    Contol_Mean
## treat                    "0"           "1"
## cohortdate               "2005-05-16"  "2005-01-12"
## Admsn_source             NA            NA
## adm_type                 NA            NA
## BENE_ID                  NA            NA
## Index_admsn_dt           "2005-05-14"  "2005-01-10"
## Index_dschrg_dt          "2005-05-23"  "2005-01-19"
## surgdate                 "2005-05-16"  "2005-01-12"
## PRVDR_NUM                NA            NA
## HospCABGVOl_1999         NA            NA
## HospCABGVOl_2000         NA            NA
## HospCABGVOl_2001         NA            NA
## HospCABGVOl_2002         NA            NA
## HospCABGVOl_2003         NA            NA
## HospCABGVOl_2004         NA            NA
## HospCABGVOl_2005         NA            NA
## HospCABGVOl_2006         NA            NA
## HospCABGVOl_2007         NA            NA
## HospCABGVOl_2008         NA            NA
## HospCABGVOl_2009         NA            NA
## HospCABGVOl_2010         NA            NA
## HospCABG_total           "2930.559"    "2939.781"
## HospCABG_mean            "249.2153"    "249.4890"
## hospname                 NA            NA
```

```
## zipcode                    NA            NA
## hosp_type                  NA            NA
## urgency                    "0.7812179"   "0.7172588"
## disStatus                  NA            NA
## Destin                     NA            NA
## hosp_state                 NA            NA
## indexSurgyear              "2004.895"    "2004.558"
## deathdt                    NA            NA
## CheckDeath                 NA            NA
## covend                     "2010-02-20"  "2010-03-18"
## dob                        "1933-02-23"  "1933-12-26"
## COVSTART                   "1995-11-18"  "1996-05-17"
## BENE_AGE_AT_END_REF_YR     "76.36114"    "75.60406"
## BENE_RACE_CD               NA            NA
## RTI_RACE_CD                NA            NA
## BENE_PTA_TRMNTN_CD         NA            NA
## BENE_PTB_TRMNTN_CD         NA            NA
## Male                       "0.6790589"   "0.7290717"
## Sex                        NA            NA
## race                       NA            NA
## trmPartA                   NA            NA
## TrmPartB                   NA            NA
## bene_entlmt_rsn_orig       NA            NA
## bene_entlmt_rsn_curr       NA            NA
## BENE_ZIP_CD                NA            NA
## BENE_ESRD_IND              NA            NA
## BENE_MDCR_STATUS_CD        NA            NA
## BENE_HI_CVRAGE_TOT_MONS    NA            NA
## BENE_SMI_CVRAGE_TOT_MONS   NA            NA
## BENE_STATE_BUYIN_TOT_MONS  NA            NA
## BENE_HMO_CVRAGE_TOT_MONS   NA            NA
## dead                       "0.2668215"   "0.2408784"
## dyear                      NA            NA
## TimeTodeath_days           "1693.959"    "1847.903"
## timeTodeath_yrs            "4.637801"    "5.059297"
## ageAtdeath                 NA            NA
## ageatSurgery               "72.22462"    "71.04712"
## vein_ind                   " 1"          NA
## artery_ind                 NA            " 1"
## RadialArtery_ind           NA            NA
## veinct                     "2.469851"    "1.708752"
## arteryCt                   "0.9986069"   "2.0000000"
## totalct                    "3.468458"    "3.708752"
## CABGnpi1                   NA            NA
## CABGmeanVol1               NA            NA
## CABGtotalVol1              NA            NA
## CABGmdname1                NA            NA
## CABGwebSpec1               NA            NA
## CABGspeccode1              NA            NA
## cabg_Othtotal1             NA            NA
## cabg_OthMean1              NA            NA
## AMI_7                      "0.07465450"  "0.06983025"
## ALZH_7                     "0.011032287" "0.008417997"
## ALZH_DEMEN_7               "0.03911339"  "0.03134978"
```

```
## ATRIAL_FIB_7              "0.07074290"  "0.06574115"
## CATARACT_7                "0.4988521"   "0.4566290"
## CHRONICKIDNEY_7           "0.1408257"   "0.1153404"
## COPD_7                    "0.2069664"   "0.1758440"
## CHF_7                     "0.2494214"   "0.2223765"
## DIABETES_7                "0.4151271"   "0.3851455"
## GLAUCOMA_7                "0.1479636"   "0.1378557"
## HIP_FRACTURE_7            "0.009486733" "0.007256894"
## ISCHEMICHEART_7           "0.7015238"   "0.7061652"
## DEPRESSION_7              "0.1500301"   "0.1408973"
## OSTEOPOROSIS_7            "0.06922253"  "0.05506405"
## RA_OA_7                   "0.3539465"   "0.3277213"
## STROKE_TIA_7              "0.1205188"   "0.1042721"
## CANCER_BREAST_7           "0.01685264"  "0.01416041"
## CANCER_COLORECTAL_7       "0.02140579"  "0.01851455"
## CANCER_PROSTATE_7         "0.06685648"  "0.06562756"
## CANCER_LUNG_7             "0.007307581" "0.005729791"
## CANCER_ENDOMETRIAL_7      "0.002868421" "0.002498896"
## ANEMIA_7                  "0.3472698"   "0.3073011"
## ASTHMA_7                  "0.07162304"  "0.06386067"
## HYPERL_7                  "0.6875462"   "0.6787909"
## HYPERP_7                  "0.1984645"   "0.1972108"
## HYPERT_7                  "0.7756759"   "0.7459077"
## HYPOTH_7                  "0.10364093"  "0.09109611"
## Ami_Index                 "0.1492971"   "0.1397993"
## cancerhistory_7           "0.1102221"   "0.1024421"
## revasc_Date               NA            NA
## Stroke_date               NA            NA
## bYPASS361_date            NA            NA
## bYPASS361_1               NA            NA
## pci_date                  NA            NA
## pci_1                     NA            NA
## ACUTEmi_date              NA            NA
## ACUTEmi_1                 NA            NA
## revasc_Event              NA            NA
## stratum                   "539.8588"    "563.1397"
##
## attr(,"class")
## [1] "summary.strata"
```

```
plot(m.strat)
```


1.00
0.75
0.50
0.25
0.00
Fraction Control Observations
0
10000
20000
30000
40000
Stratum Size

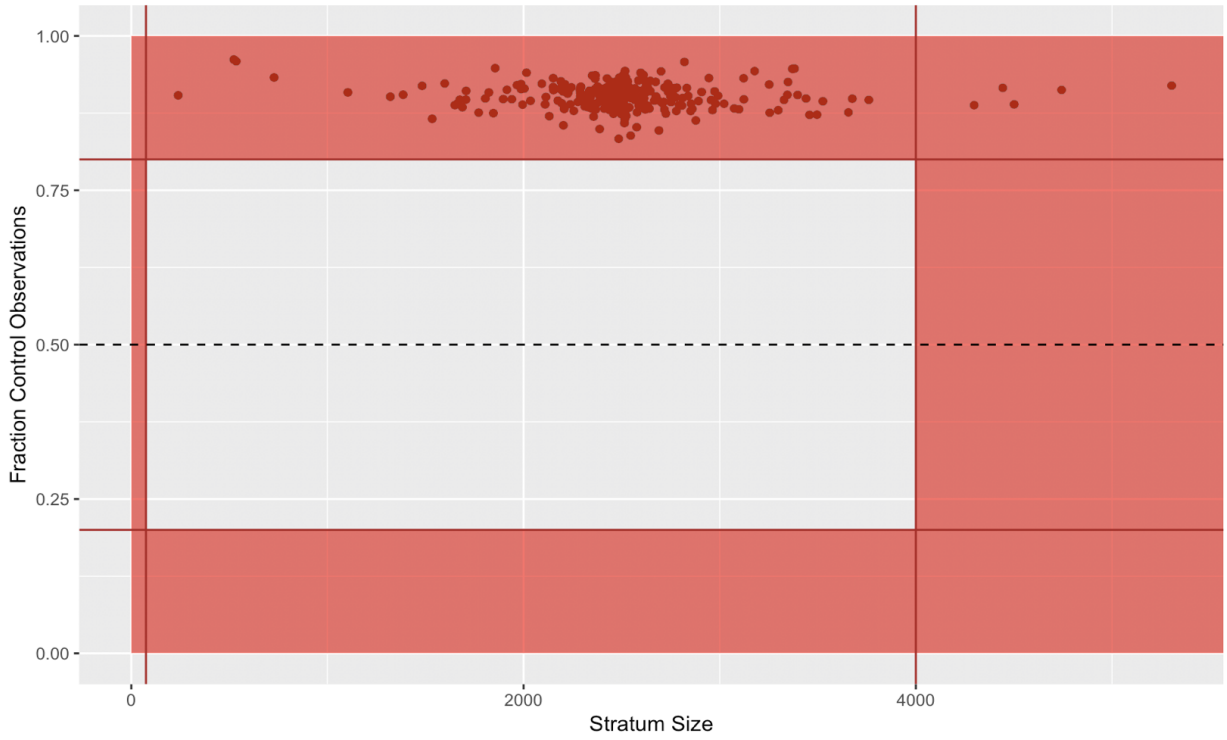


Figure 1: Auto Stratify plot.

Auto Stratify by Prognostic Score

```
a.strat <- auto_stratify(data = dat, treat = "treat",
                         outcome = "dead",
                         covariates = c("totalct", "hosp_state", "AMI_7", "COPD_7",
                                        "ISCHEMICHEART_7","STROKE_TIA_7", "ATRIAL_FIB_7",
                                        "CHRONICKIDNEY_7", "DIABETES_7","ALZH_DEMEN_7",
                                        "Male", "race"))
```

```
plot(a.strat)
```