A Short Demo for the Anonymization Procedure

Sinan Shi

October 28, 2016

```
library(ccdata)
library(ccanonym)
```

The YAML configuration file

Create a YAML configuration file as such, where the **identifiable variables** (dirctVars), **key categorical variables** (key-Vars), **key numerical variables** (numVars), **key date-time variables** (datetimeVars), **sensitive variables** (sensVars) and the corresponding operations and thresholds are specified.

```
k-anonymity: 30
l-diversity: 30


directVars:
    - pasno      # PAS number
    - ICNNO      # Site code
    - ADNO       # INCNARC admission number
    - NHSNO      # NHS number
    - TUADNO     # Transferring unit admission number
    - DOB          # Date of birth

categoricalVars:
    - GPCODE     # GP code
    - SEX          # Sex
    - PCODE      # Postcode

sensVars:
    - BPC        # Biopsy proven cirrhosis
    - AIDS_V3 # HIC/AIDS
    - PH           # Portal hypertension
    - RAICU1     # Primary reason for admission to your unit
    - RAICU2     # Secondary reasons for admission to your unit
    - URAICU     # Ultimate primary reason for admission to unit

numVars:
    HCM: # Height
        microaggregation:
            aggr: 2
```

```
datetimeVars:
    DOAH:   # Date of original admission to/attendance at acute hospital'
        microaggregation:
            aggr: 1
        addNoise:
            noise: 2

    DAH: # Date of admission to your hospital
        addNoise:
            noise: 2

    DOAICU:  # Date of original admission to ICU/HDU'
        microaggregation:
            aggr: 1
        addNoise:
            noise: 1
... ...
```

```r
conf <- yaml.load_file("../data/test_demo.yaml")
vars <- anony.var(conf)
```

Identifiable data set

The identifiable data set is usually stored in ccRecord format. In the following code, we create the ccRecord object from a XML file which contains only five episodes.

```r
ccd <- xml2Data("../tests/data/test_data_anonym.xml")
demg.table <- as.data.frame(sql.demographic.table(ccd))
print(conf)
```

```r
all.var <- c(vars$dirv, vars$all.vars, "DIS") # all variables besides non-confidential data
pander(demg.table[, all.var],  style = 'rmarkdown')
```

pasno	ICNNO	NHSNO	DOB	GPCODE	SEX	PCODE	HCM	DAH	AIDS_V3	DIS
pas_1	site_1	nhs_1	1988-06-07	GPCODE1	F	NW1 1BB	192	2014-02-01	TRUE	D
pas_2	site_1	nhs_2	1980-12-30	GPCODE1	F	NW1 1BB	174	2014-02-01	FALSE	D
pas_1	site_1	nhs_1	1970-01-25	GPCODE2	F	NW1 1BB	170	2014-02-09	FALSE	D
pas_1	site_1	nhs_1	1977-01-25	GPCODE2	M	NW1 2BB	160	2014-02-01	FALSE	D
pas_1	site_1	nhs_1	1955-01-04	GPCODE4	M	NULL	165	2014-02-01	FALSE	D
pas_1	site_1	nhs_1	1988-06-07	GPCODE1	F	NW1 1BB	170	2014-02-01	TRUE	L
pas_1	site_1	nhs_1	1988-06-07	GPCODE1	F	NW1 1BB	170	2014-02-01	TRUE	NULL

Anonymisation

Anonymisation procedure

```
anonccd <- anonymisation(ccd, conf, remove.alive=T)
demg <- data.frame(sql.demographic.table(anonccd))
pander(demg[, all.var], style="rmarkdown")
```

Table 0.2: Table continues below

pasno	ICNNO	NHSNO	DOB	GPCODE	SEX	PCODE	HCM	DAH
NULL	NULL	NULL	NULL	GPCODE1	F	NW1 1BB	178.7	2014-02-05 08:25:48
NULL	NULL	NULL	NULL	GPCODE1	F	NW1 1BB	178.7	2014-02-03 03:08:10
NULL	NULL	NULL	NULL	NULL	F	NW1 1BB	178.7	2014-02-07 15:37:26
NULL	NULL	NULL	NULL	NULL	M	NW1 2BB	162.5	2014-02-02 13:13:44
NULL	NULL	NULL	NULL	GPCODE4	M	NULL	162.5	2014-01-31 13:56:35

AIDS_V3	DIS
TRUE	D
FALSE	D
FALSE	D
FALSE	D
FALSE	D

Categorical variables

```
print(conf)
```

```
## $directVars
## [1] "pasno" "ICNNO" "NHSNO" "DOB"
##
## $keyVars
## [1] "GPCODE" "SEX"     "PCODE"
##
## $sensVars
## [1] "AIDS_V3"
##
## $numVars
## $numVars$HCM
## $numVars$HCM$microaggregation
## $numVars$HCM$microaggregation$aggr
## [1] 2
##
##
##
##
```

```
## $datetimeVars
## $datetimeVars$DAH
## $datetimeVars$DAH$addNoise
## $datetimeVars$DAH$addNoise$noise
## [1] 100
##
##
##
##
## $nonidentifyVars
##    [1] "DUDICU"              "DWFRD"
##    [3] "TWFRD"               "DDH"
##    [5] "DAICU"               "DDICU"
##    [7] "DDBSD"               "DOAICU"
##    [9] "ADNO"                "DUDH"
##   [11] "DOAH"                "TUADNO"
##   [13] "DLCCA"               "RAICU1"
##   [15] "RAICU2"              "URAICU"
##   [17] "BPC"                 "PH"
##   [19] "DOD"                 "SCODE"
##   [21] "CCL2D"               "ORGAN_SUPPORT"
##   [23] "AMLALLMM"            "PSP"
##   [25] "TGA"                 "TGD"
##   [27] "AMUAI"               "HCMEST"
##   [29] "WKG"                 "WKGEST"
##   [31] "BCSD"                "ACSD"
##   [33] "CHEMOX"              "CMLCLL"
##   [35] "CRRX"                "CLASSNS"
##   [37] "CICIDS"              "CCD"
##   [39] "CCA"                 "DBRICU"
##   [41] "TBRICU"              "TOD"
##   [43] "TDBSD"               "DTW"
##   [45] "TTW"                 "TNESSA"
##   [47] "DHRS"                "DEP"
##   [49] "DSD"                 "HLOCD"
##   [51] "ETHNIC"              "GSD"
##   [53] "HE"                  "HV"
##   [55] "HLOCA"               "LEVD"
##   [57] "LSD"                 "LOCA"
##   [59] "LOCD"                "LYM"
##   [61] "META"                "NSD"
##   [63] "OCPMH"               "PLOCA"
##   [65] "RADIOX"              "REFOD"
##   [67] "RSD"                 "RESD"
##   [69] "RESA"                "BRSD"
##   [71] "ARSD"                "SOHA"
##   [73] "SOHD"                "SRD"
##   [75] "HDIS"                "DIS"
##   [77] "UHDIS"               "STERX"
##   [79] "TUIDI"               "TYPEIHA"
##   [81] "VSCD"                "h_rate"
##   [83] "h_rhythm"            "bp_m_a"
##   [85] "bp_m_ni"             "bp_sys_a"
##   [87] "bp_sys_ni"           "bp_dia_a"
```

```
##   [89] "bp_dia_ni"                 "venous_p"
##   [91] "lidco_plus"                "lidco_rapid"
##   [93] "picco"                     "pa_catheter"
##   [95] "doppler"                   "lactate_abg"
##   [97] "lactate_lab"               "venous_saturation"
##   [99] "airway"                    "spo2"
##  [101] "sao2_abg"                  "pao2_abg"
##  [103] "paco2_abg"                 "ph_abg_vbg"
##  [105] "temperature_central"       "temperature_non_central"
##  [107] "position"                  "ventilation"
##  [109] "respiratory_rate_totl"     "total_resp_rate"
##  [111] "mand_resp_rate"            "minute_volume"
##  [113] "airway_pressure"           "fraction_oxygen"
##  [115] "pe_expiratory_pressure"    "airway_pressure"
##  [117] "frequency"                 "cycle_volumn"
##  [119] "base_flow"                 "gcs_total"
##  [121] "gcs_motor"                 "gcs_eye"
##  [123] "gcs_verbal"                "sedation_score"
##  [125] "renal_replace"             "urine_output"
##  [127] "urea"                      "creatinine"
##  [129] "sodium"                    "sodium_abg_vbg"
##  [131] "potassium"                 "potassium_abg_vbg"
##  [133] "bilirubin"                 "glucode_abg_vbg"
##  [135] "glucode_bedtest"           "haemoglobin_abg_vbg"
##  [137] "haemoglobin"               "white_cell"
##  [139] "neutrophil"                "platelets"
##  [141] "site"                      "organism"
##  [143] "sensitivity"               "fentanyl"
##  [145] "milrinone"                 "bed05"
##  [147] "bed50"                     "PA_V3"
##  [149] "BSDTP"                     "TNESSD"
##  [151] "DFCCD"                     "apache_score"
##  [153] "apache_prob"               "fluid_balance"
##  [155] "amikacin"                  "amoxicillin"
##  [157] "azithromycin"              "benzylpenicillin"
##  [159] "cefotaxime"                "ceftazidime"
##  [161] "ceftriaxone"               "cefuroxime"
##  [163] "chloramphenicol"           "ciprofloxacin"
##  [165] "clarithromycin"            "clindamycin"
##  [167] "co-amoxiclav"              "colistin"
##  [169] "co-trimoxazole"            "demeclocycline hcl"
##  [171] "doxycycline"               "ertapenem"
##  [173] "erythromycin"              "ethambutal hcl"
##  [175] "flucloxacillin"            "fuscidic acid"
##  [177] "gentamicin"                "isoniazid"
##  [179] "levofloxacin"              "linezolid"
##  [181] "meropenem"                 "metronidazole"
##  [183] "moxifloxacin"              "neomycin"
##  [185] "nitrofurantion"            "ofloxacin"
##  [187] "pentamidine"               "phenoxymethylpenicillin"
##  [189] "piperacillin_tazobactam"   "pyrazinamide"
##  [191] "rifampacin"                "rifater"
##  [193] "rifinah"                   "sodium fusidate"
##  [195] "teicoplanin"               "tigecycline"
```

```
## [197] "tobramycin"          "trimethoprim"
## [199] "vancomycin"          "propofol"
## [201] "midazolam"           "remifentanil"
## [203] "adrenaline"          "dobutamine"
## [205] "dopamine"            "enoximone"
## [207] "levosimendan"        "noradrenaline"
## [209] "vasopressin"         "respiratory_rate_spon"
## [211] "tidal_volume"        "duration_therapy"
## [213] "effluent_per_day"    "dialysate"
## [215] "replace_fluid_RRT"   "anticoagulation"
## [217] "cprotein"            "thiopentone_thiopental"
## [219] "clonidine"           "dexmedetomidine"
## [221] "ketamine"            "morphine"
## [223] "dopexamine"          "terlipressin"
## [225] "TYPEIHD"             "esmolol"
## [227] "metoprolol"          "dexamethasone"
## [229] "hydrocortisone"      "methylprednisolone"
## [231] "sedataion"           "pao2_fio2"
## [233] "fluid_balance"       "glucose_lab"
## [235] "UDIS"                "advsupt_resp"
## [237] "basicsupt_resp"      "advsupt_resp"
## [239] "basicsupt_cardv"     "supt_renal"
## [241] "supt_neuro"          "supt_liver"
## [243] "supt_dermat"         "supt_gastr"
## [245] "bed02"               "bed03"
## [247] "CCL3D"               "RDIS_V3"
## [249] "DESTH_V3"            "CPR_V3"
## [251] "ITW_V3"              "OD_V3"
## [253] "hco3_abg_vbg"
```