```
 1 . /**********************************************\
   > * Practice with Stata: data downloaded from internet***/
 2 . * Date modified:
 3 . * Output saved in: "/Users/Banjodan2/Desktop/StataPractice/"
 4 .
 5 . label define binaryCode 0 "No" 1 "Yes"

 6 .
 7 . import delimited datasets/KaggleV2-May-2016.csv //hospital no-show rates
   (14 vars, 110527 obs)

 8 . //use
 9 .
10 .
11 . //clean the data:
12 . drop if age < 0
   (1 observation deleted)

13 . foreach var of varlist noshow {
     2.          encode `var', gen(_`var')
     3.          replace _`var' = _`var' - 1
     4.          label values _`var' binaryCode
     5. }
   (110526 real changes made)

14 .
15 . foreach var of varlist scholarship hipertension diabetes alcoholism handcap
   > sms_received {
     2.          label values `var' binaryCode
     3. }

16 .
17 . numlabel, add

18 .
19 . local categoricalVars "gender scholarship hipertension diabetes alcoholism h
   > andcap sms_received noshow neighbourhood"
```

```
20 . local size : word count `categoricalVars'

21 .
22 . di "We have `size' categorical variables."
   We have 9 categorical variables.

23 .
24 . *summarizing the categorical vars:
25 . foreach var of varlist `categoricalVars' {
   2.           sum `var'
   3.           tabulate `var', missing plot
   4. }
```

Variable	Obs	Mean	Std. Dev.	Min	Max
gender	0				

Gender	Freq.

```
> ─
       F |    71,839 |**************************************************
> *
       M |    38,687 |****************************
```

Total	110,526

Variable	Obs	Mean	Std. Dev.	Min	Max
scholarship	110526	.0982665	.297676	0	1

Scholarship	Freq.

```
> ─
    0. No |    99,665 |**************************************************
> *
   1. Yes |    10,861 |******
```

Total	110,526

Variable	Obs	Mean	Std. Dev.	Min	Max
hipertension	110526	.1972477	.3979227	0	1

STaTa®

```
Hipertensio
          n |        Freq.
------------+------------------
> --
     0. No |       88,725 |*********************************************
> *
     1. Yes |      21,801 |*************
------------+------------------
> --
       Total |      110,526
```

Variable	Obs	Mean	Std. Dev.	Min	Max
diabetes	110526	.0718654	.2582662	0	1

```
  Diabetes |        Freq.
------------+------------------
> --
     0. No |      102,583 |**********************************************
> *
     1. Yes |       7,943 |****
------------+------------------
> --
       Total |      110,526
```

Variable	Obs	Mean	Std. Dev.	Min	Max
alcoholism	110526	.0304001	.1716863	0	1

```
 Alcoholism |        Freq.
------------+------------------
> --
     0. No |      107,166 |**********************************************
> *
     1. Yes |       3,360 |**
------------+------------------
> --
       Total |      110,526
```

Variable	Obs	Mean	Std. Dev.	Min	Max
handcap	110526	.0222482	.1615434	0	4

```
       Handcap |      Freq.
---------------+---------------------------------------------------------------
> ─
         0. No |    108,285 |*****************************************************
> *
        1. Yes |      2,042 |*
             2 |        183
             3 |         13
             4 |          3
---------------+---------------------------------------------------------------
> ─
         Total |    110,526
```

```
    Variable |        Obs        Mean    Std. Dev.       Min        Max
-------------+--------------------------------------------------------------
sms_received |    110526    .3210285    .4668738          0          1
```

```
   SMS_receive |
             d |      Freq.
---------------+---------------------------------------------------------------
> ─
         0. No |     75,044 |*****************************************************
> *
        1. Yes |     35,482 |***********************
---------------+---------------------------------------------------------------
> ─
         Total |    110,526
```

```
    Variable |        Obs        Mean    Std. Dev.       Min        Max
-------------+--------------------------------------------------------------
      noshow |          0
```

```
    No-show |      Freq.
-----------+---------------------------------------------------------------
> ─
        No |     88,207 |*******************************************************
> *
       Yes |     22,319 |*************
-----------+---------------------------------------------------------------
> ─
     Total |    110,526
```

```
    Variable |        Obs        Mean    Std. Dev.       Min        Max
-------------+--------------------------------------------------------------
 neighbourh~d |          0
```

```
                 Neighbourhood │     Freq. │
─────────────────────────────────┼───────────┼──────────────────────
> ─
                     AEROPORTO │         8 │
                    ANDORINHAS │     2,262 │***********
               ANTÃ?NIO HONÃ?RIO │       271 │*
              ARIOVALDO FAVALESSA │       282 │*
                 BARRO VERMELHO │       423 │**
                    BELA VISTA │     1,907 │*********
                BENTO FERREIRA │       858 │****
                     BOA VISTA │       312 │*
                        BONFIM │     2,773 │*************
                   CARATOÃ?RA │     2,565 │************
                        CENTRO │     3,334 │****************
                       COMDUSA │       310 │*
                      CONQUISTA │       849 │****
                  CONSOLAÃ?Ã?O │     1,376 │******
                    CRUZAMENTO │     1,398 │*******
                      DA PENHA │     2,217 │**********
                     DE LOURDES │       305 │*
                      DO CABRAL │       560 │***
                     DO MOSCOSO │       413 │**
                      DO QUADRO │       849 │****
                 ENSEADA DO SUÃ? │       235 │*
                    ESTRELINHA │       538 │***
                   FONTE GRANDE │       682 │***
               FORTE SÃ?O JOÃ?O │     1,889 │*********
                     FRADINHOS │       258 │*
                    GOIABEIRAS │       700 │***
                GRANDE VITÃ?RIA │     1,071 │*****
                      GURIGICA │     2,018 │*********
                        HORTO │       175 │*
              ILHA DAS CAIEIRAS │     1,071 │*****
            ILHA DE SANTA MARIA │     1,885 │*********
                   ILHA DO BOI │        35 │
                  ILHA DO FRADE │        10 │
               ILHA DO PRÃ?NCIPE │     2,266 │***********
    ILHAS OCEÃ?NICAS DE TRINDADE │         2 │
                  INHANGUETÃ? │     1,057 │*****
                      ITARARÃ? │     3,514 │****************
                       JABOUR │     2,509 │************
                 JARDIM CAMBURI │     7,717 │*********************************
> *
                JARDIM DA PENHA │     3,877 │******************
             JESUS DE NAZARETH │     2,853 │*************
                 JOANA DÂ´ARC │     1,427 │*******
                  JUCUTUQUARA │       694 │***
                   MARIA ORTIZ │     5,805 │*************************
                    MARUÃ?PE │     1,902 │*********
```

```
        MATA DA PRAIA |      644 |***
          MONTE BELO |      824 |****
   MORADA DE CAMBURI |       96 |
      MÃ?RIO CYPRESTE |      371 |**
            NAZARETH |      135 |*
      NOVA PALESTINA |    2,264 |***********
    PARQUE INDUSTRIAL |        1 |
      PARQUE MOSCOSO |      802 |****
             PIEDADE |      452 |**
    PONTAL DE CAMBURI |       69 |
      PRAIA DO CANTO |    1,035 |*****
       PRAIA DO SUÃ? |    1,288 |******
          REDENÃ?Ã?O |    1,553 |*******
           REPÃ?BLICA |      835 |****
         RESISTÃ?NCIA |    4,431 |********************
             ROMÃ?O |    2,214 |**********
      SANTA CECÃ?LIA |      448 |**
         SANTA CLARA |      506 |**
        SANTA HELENA |      178 |*
        SANTA LUÃ?ZA |      428 |**
        SANTA LÃ?CIA |      438 |**
       SANTA MARTHA |    3,131 |***************
        SANTA TEREZA |    1,332 |******
        SANTO ANDRÃ? |    2,571 |************
      SANTO ANTÃ?NIO |    2,746 |*************
       SANTOS DUMONT |    1,276 |******
         SANTOS REIS |      547 |***
    SEGURANÃ?A DO LAR |      145 |*
        SOLON BORGES |      469 |**
        SÃ?O BENEDITO |    1,439 |*******
     SÃ?O CRISTÃ?VÃ?O |    1,836 |*********
          SÃ?O JOSÃ? |    1,977 |*********
          SÃ?O PEDRO |    2,448 |**********
          TABUAZEIRO |    3,132 |**************
      UNIVERSITÃ?RIO |      152 |*
          VILA RUBIM |      851 |****
─────────────────────┼──────────┼──────────────────────────────
                     |          |
> ─
             Total |  110,526
```

```
26 .
27 . notes gender: there is no significant difference between the no-show rates o
  > f males and females

28 .
29 . save "PracticeDataCleaned.dta", replace
  file PracticeDataCleaned.dta saved

30 .
31 . *tabbout eststo esttab estadd
32 . eststo: regress _noshow sms_received age
```

Source	SS	df	MS		Number of obs	=	110526
					F(2,110523)	=	1117.39
Model	353.020515	2	176.510258		Prob > F	=	0.0000
Residual	17459.0059	110523	.157967173		R-squared	=	0.0198
					Adj R-squared	=	0.0198
Total	17812.0264	110525	.161158348		Root MSE	=	.39745

_noshow	Coef.	Std. Err.	t	P>|t|	[95% Conf. Interval]	
sms_received	.1093829	.0025609	42.71	0.000	.1043636	.1144022
age	-.0010759	.0000517	-20.80	0.000	-.0011773	-.0009745
_cons	.206722	.0023973	86.23	0.000	.2020233	.2114208

```
  (est2 stored)

33 . esttab using "output/tables/table1_regress_noshow_sms.tex", replace se r2 no
  > cons booktabs
  (output written to output/tables/table1_regress_noshow_sms.tex)

34 . eststo clear

35 .
36 . //Great way to turn a string of numbers into integer values and remove unwan
  > ted characters from a variable:
```

```
37 . * gen var2=regexr(var1,"[.\}\)\*a-zA-Z]+","")
38 . * destring var2, replace
39 .
40 . * or to extract strings:
41 . * gen var2=regexr(var1,"[.0-9]+","")
42 .
43 . * 1. -by:-.
44 .
45 . * 2. -foreach- or -forval- with varlists or numlists.
46 .
47 . * 3. -merge-. I rarely use it but -merge-masters have real leverage
48 . * in file manipulations.
49 .
50 . * 4. -assert-. My candidate for the most underestimated command in
51 . * Stata (second is -count-).
52 .
53 . * 5. -reshape-.
54 .
55 . *other stata commands to try:
56 . /* margins
   > svtmat
   > coeffplot
   > collapse */
57 .
58 . //snapshot or preserve
```

STATA®