```
-------------------------------------------------------------------------------
-------------------------------------------------------------------------------
----------
      name:  <unnamed>
  log type:  text
 opened on:  18 Sep 2019, 15:53:07

. /* ====================================================================
>                                          1. Merge Regular Firms Data
>  ==================================================================== */
> * Drop redudant variables and merge the files ;
. local x = 1 ;

. forvalues x = 1/4{;
  2.         use reg`x'.dta, clear ;
  3.         drop chr* ;
  4.         drop dm2* ;
  5.         drop dm3* ;
  6.         drop dm1_hours ;
  7.         drop dm1_ef ;
  8.         save reg`x'.dta, replace ;
  9. };
file reg1.dta saved
file reg2.dta saved
file reg3.dta saved
file reg4.dta saved

. use reg1.dta, clear ;

. forvalues x = 2/4{;
  2.         append using reg`x'.dta ;
  3. };
firm_id was str25 now str26
type was int now double

. save regall.dta, replace ;
(note: file regall.dta not found)
file regall.dta saved

. * Merge Regular with Key Firms
> clear ;
. use keynum.dta, clear ;

. * Label Key Firms Census_Type = 1, Regular firms = 2 ;
. generate Census_Type = 1 ;

. save keynum.dta, replace ;
file keynum.dta saved

. use regall.dta, clear ;
```

```
. generate Census_Type = 2 ;

.  save regall.dta, replace ;
file regall.dta saved

. append using keynum.dta ;

. save allfirms.dta, replace ;
(note: file allfirms.dta not found)
file allfirms.dta saved

. /* ================================================================
>                                                  2. Labeling Variables
>  ================================================================ */
> * ----------------- Key Firms ----------------------;
. clear ;

. use keynum.dta, clear ;

. * Label Key Firms Variables ;
. label variable firm_id "Firm ID" ;

. label variable areacode "Area Code up to the Level of Subdistrict";

. label variable industry "4-digits Industrial Sector Code (GB2002)";

. label variable type "Ownership Rights of the Firm";

. label variable opr_hours "Total Annual Operating Hours";

.  label variable product "Total VALUE of Output (in RMB10000)";

. label variable quantity1 "Total QUANTITY of Output 1";

. label variable quantity2 "Total QUANTITY of Output 2";

. label variable quantity3 "Total QUANTITY of Output 3";

. label variable quantity4 "Total QUANTITY of Output 4";

. label variable quantity5 "Total QUANTITY of Output 5";

. label variable wastewater_g ///
>         "Total Amount of Wastewater Generated (in ton)";

. label variable wastewater_e ///
>         "Total Amount of Wastewater Discharged (in ton)";

. label variable water_u ///
```

```
>         "Total Water Used = Acquired + Recycled (in ton)";

. label variable water_r ///
>         "Total Water Recycled (in ton)";

. label variable cod_g ///
>         "Total Amount of Chemical Oxygen Demand Generated (in ton)";

. label variable cod_e ///
>         "Total Amount of Chemical Oxygen Demand Discharged (in ton)";

. label variable nh_g ///
>         "Total Amount of Ammonian Generated (in ton)";

. label variable nh_e ///
>         "Total Amount of Ammonian Discharged (in ton)";

. label variable pet_g ///
>         "Total Amount of Petroleum Generated (in ton)";

. label variable pet_e ///
>         "Total Amount of Petroleum Discharged (in ton)";

.  label variable phe_g ///
>         "Total Amount of Volatile Phenol Generated (in ton)";

. label variable phe_e ///
>         "Total Amount of Volatile Phenol Discharged (in ton)";

. label variable bod_g ///
>         "Total Amount of Biochemical Oxygen Demand Generated (in ton)";

. label variable bod_e ///
>         "Total Amount of Biochemical Oxygen Demand Discharged (in ton)";

. label variable cyn_g ///
>         "Total Amount of Cyanidum Generated (in kg)";

. label variable cyn_e ///
>         "Total Amount of Cyanidum Discharged (in kg)";

. label variable as_g ///
>         "Total Amount of Arsenium Generated (in kg)";

. label variable as_e ///
>         "Total Amount of Arsenium Discharged (in kg)";

. label variable chr_g ///
>         "Total Amount of Chromium Generated (in kg)";
```

```
. label variable chr_e ///
>         "Total Amount of Chromium Discharged (in kg)";

. label variable chr6_g ///
>         "Total Amount of Hexavalent Chrome Generated (in kg)";

. label variable chr6_e ///
>         "Total Amount of Hexavalent Chrome Discharged (in kg)";

. label variable dm1_inv ///
>         "Total Investment of Wastewater Disposal Equipment 1 (RMB 10000)";

. label variable dm1_quant ///
>         "Designed Disposal Capacity of Equip 1 (in ton)";

. label variable dm1_oprcost ///
>         "Total Operating Costs of Equip 1 (RMB 10000)";

. label variable dm1_hours ///
>         "Total Operating Hours of Equip 1";

. label variable dm1_elec ///
>         "Total Amount of Electricity Consumed Equip 1 (10000kw/h)";

. label variable dm1_ef ///
>         "Efficiency of Equip 1 (in %)";

. label variable dm1_code ///
>         "Code for Disposal Method of Equip 1";

. label variable dm2_inv ///
>         "Total Investment of Wastewater Disposal Equipment 2 (RMB 10000)";

. label variable dm2_quant ///
>         "Designed Disposal Capacity of Equip 2 (in ton)";

. label variable dm2_oprcost ///
>         "Total Operating Costs of Equip 2 (RMB 10000)";

. label variable dm2_hours ///
>         "Total Operating Hours of Equip 2";

. label variable dm2_elec ///
>         "Total Amount of Electricity Consumed Equip 2 (10000kw/h)";

. label variable dm2_ef ///
>         "Efficiency of Equip 2 (in %)";

. label variable dm2_code ///
>         "Code for Disposal Method of Equip 2";
```

```
. label variable dm3_inv ///
>         "Total Investment of Wastewater Disposal Equipment 3 (RMB 10000)";

. label variable dm3_quant ///
>         "Designed Disposal Capacity of Equip 3 (in ton)";

. label variable dm3_oprcost ///
>         "Total Operating Costs of Equip 3 (RMB 10000)";

. label variable dm3_hours ///
>         "Total Operating Hours of Equip 3";

. label variable dm3_elec ///
>         "Total Amount of Electricity Consumed Equip 3 (10000kw/h)";

. label variable dm3_ef ///
>         "Efficiency of Equip 3 (in %)";

. label variable dm3_code ///
>         "Code for Disposal Method of Equip 3";

. label variable Census_Type ///
>         "Census Type Code: 1 Key 2 Regular";

. * Aggregate counties to provinces ;
. recode areacode (110000000000/119999999999 = 11)
>         (120000000000/129999999999 = 12) (130000000000/139999999999 = 13)
>         (140000000000/149999999999 = 14) (150000000000/159999999999 = 15)
>         (210000000000/219999999999 = 21) (220000000000/229999999999 = 22)
>         (230000000000/239999999999 = 23) (310000000000/319999999999 = 31)
>         (320000000000/329999999999 = 32) (330000000000/339999999999 = 33)
>         (340000000000/349999999999 = 34) (350000000000/359999999999 = 35)
>         (360000000000/369999999999 = 36) (370000000000/379999999999 = 37)
>         (410000000000/419999999999 = 41) (420000000000/429999999999 = 42)
>         (430000000000/439999999999 = 43) (440000000000/449999999999 = 44)
>         (450000000000/459999999999 = 45) (460000000000/469999999999 = 46)
>         (500000000000/509999999999 = 50) (510000000000/519999999999 = 51)
>         (520000000000/529999999999 = 52) (530000000000/539999999999 = 53)
>         (540000000000/549999999999 = 54) (610000000000/619999999999 = 61)
>         (620000000000/629999999999 = 62) (630000000000/639999999999 = 63)
>         (640000000000/649999999999 = 64) (650000000000/659999999999 = 65),
>         generate(province) ;
(106067 differences between areacode and province)

. * Aggregate sectors ;
. recode industry (500/599 = 5)
>         (600/699 = 6) (700/799 = 7) (800/899 = 8) (900/999 = 9)
>         (1000/1099 = 10) (1100/1199 = 11) (1300/1399 = 13) (1400/1499 = 14)
>         (1500/1599 = 15) (1600/1699 = 16) (1700/1799 = 17) (1800/1899 = 18)
```

```
>         (1900/1999 = 19) (2000/2099 = 20) (2100/2199 = 21) (2200/2299 = 22)
>         (2300/2399 = 23) (2400/2499 = 24) (2500/2599 = 25) (2600/2699 = 26)
>         (2700/2799 = 27) (2800/2899 = 28) (2900/2999 = 29) (3000/3099 = 30)
>         (3100/3119 = 31) (3200/3299 = 32) (3300/3399 = 33) (3400/3499 = 34)
>         (3500/3599 = 35) (3600/3699 = 36) (3700/3799 = 37) (3900/3999 = 39)
>         (4000/4099 = 40) (4100/4199 = 41) (4200/4299 = 42) (4300/4399 = 43)
>         (4400/4499 = 44) (4500/4599 = 45) (4600/4699 = 46),
>         generate(industry_a) ;
(84965 differences between industry and industry_a)

. save keynum.dta, replace ;
file keynum.dta saved

.         * ----------------- Regular Firms ----------------------;
. clear ;

. use regall.dta, clear ;

. * Label Regular Firms variables ;
. label variable firm_id "Firm ID" ;

. label variable areacode "Area Code up to the Level of Subdistrict";

. label variable industry "4-digits Industrial Sector Code (GB2002)";

. label variable type "Ownership Rights of the Firm";

. label variable opr_hours "Total Annual Operating Hours";

.  label variable product "Total VALUE of Output (in RMB10000)";

. label variable quantity1 "Total QUANTITY of Output 1";

. label variable quantity2 "Total QUANTITY of Output 2";

. label variable quantity3 "Total QUANTITY of Output 3";

. label variable quantity4 "Total QUANTITY of Output 4";

. label variable quantity5 "Total QUANTITY of Output 5";

. label variable wastewater_g ///
>         "Total Amount of Wastewater Generated (in ton)";

. label variable wastewater_e ///
>         "Total Amount of Wastewater Discharged (in ton)";

. label variable water_u ///
>         "Total Water Used = Acquired + Recycled (in ton)";
```

```
. label variable water_r ///
>         "Total Water Recycled (in ton)";

. label variable cod_g ///
>         "Total Amount of Chemical Oxygen Demand Generated (in ton)";

. label variable cod_e ///
>         "Total Amount of Chemical Oxygen Demand Discharged (in ton)";

. label variable nh_g ///
>         "Total Amount of Ammonian Generated (in ton)";

. label variable nh_e ///
>         "Total Amount of Ammonian Discharged (in ton)";

. label variable pet_g ///
>         "Total Amount of Petroleum Generated (in ton)";

. label variable pet_e ///
>         "Total Amount of Petroleum Discharged (in ton)";

.  label variable phe_g ///
>         "Total Amount of Volatile Phenol Generated (in ton)";

. label variable phe_e ///
>         "Total Amount of Volatile Phenol Discharged (in ton)";

. label variable bod_g ///
>         "Total Amount of Biochemical Oxygen Demand Generated (in ton)";

. label variable bod_e ///
>         "Total Amount of Biochemical Oxygen Demand Discharged (in ton)";

. label variable cyn_g ///
>         "Total Amount of Cyanidum Generated (in kg)";

. label variable cyn_e ///
>         "Total Amount of Cyanidum Discharged (in kg)";

. label variable as_g ///
>         "Total Amount of Arsenium Generated (in kg)";

. label variable as_e ///
>         "Total Amount of Arsenium Discharged (in kg)";

. label variable dm1_inv ///
>         "Total Investment of Wastewater Disposal Equipment 1 (RMB 10000)";

. label variable dm1_quant ///
>         "Designed Disposal Capacity of Equip 1 (in ton)";
```

```
. label variable dm1_oprcost ///
>         "Total Operating Costs of Equip 1 (RMB 10000)";

. label variable dm1_elec ///
>         "Total Amount of Electricity Consumed Equip 1 (10000kw/h)";

. label variable dm1_code ///
>         "Code for Disposal Method of Equip 1";

. label variable Census_Type ///
>         "Census Type Code: 1 Key 2 Regular";

. * Aggregate counties to provinces ;
. recode areacode (110000000000/119999999999 = 11)
>         (120000000000/129999999999 = 12) (130000000000/139999999999 = 13)
>         (140000000000/149999999999 = 14) (150000000000/159999999999 = 15)
>         (210000000000/219999999999 = 21) (220000000000/229999999999 = 22)
>         (230000000000/239999999999 = 23) (310000000000/319999999999 = 31)
>         (320000000000/329999999999 = 32) (330000000000/339999999999 = 33)
>         (340000000000/349999999999 = 34) (350000000000/359999999999 = 35)
>         (360000000000/369999999999 = 36) (370000000000/379999999999 = 37)
>         (410000000000/419999999999 = 41) (420000000000/429999999999 = 42)
>         (430000000000/439999999999 = 43) (440000000000/449999999999 = 44)
>         (450000000000/459999999999 = 45) (460000000000/469999999999 = 46)
>         (500000000000/509999999999 = 50) (510000000000/519999999999 = 51)
>         (520000000000/529999999999 = 52) (530000000000/539999999999 = 53)
>         (540000000000/549999999999 = 54) (610000000000/619999999999 = 61)
>         (620000000000/629999999999 = 62) (630000000000/639999999999 = 63)
>         (640000000000/649999999999 = 64) (650000000000/659999999999 = 65),
>         generate(province) ;
(814937 differences between areacode and province)

. * Aggregate sectors ;
. recode industry (500/599 = 5)
>         (600/699 = 6) (700/799 = 7) (800/899 = 8) (900/999 = 9)
>         (1000/1099 = 10) (1100/1199 = 11) (1300/1399 = 13) (1400/1499 = 14)
>         (1500/1599 = 15) (1600/1699 = 16) (1700/1799 = 17) (1800/1899 = 18)
>         (1900/1999 = 19) (2000/2099 = 20) (2100/2199 = 21) (2200/2299 = 22)
>         (2300/2399 = 23) (2400/2499 = 24) (2500/2599 = 25) (2600/2699 = 26)
>         (2700/2799 = 27) (2800/2899 = 28) (2900/2999 = 29) (3000/3099 = 30)
>         (3100/3119 = 31) (3200/3299 = 32) (3300/3399 = 33) (3400/3499 = 34)
>         (3500/3599 = 35) (3600/3699 = 36) (3700/3799 = 37) (3900/3999 = 39)
>         (4000/4099 = 40) (4100/4199 = 41) (4200/4299 = 42) (4300/4399 = 43)
>         (4400/4499 = 44) (4500/4599 = 45) (4600/4699 = 46),
>         generate(industry_a) ;
(773086 differences between industry and industry_a)

. save regall.dta, replace ;
file regall.dta saved
```

```
. * ----------------- All Firms ----------------------;
. clear ;

. use allfirms.dta, clear ;

. * Label All Firms variables ;
. label variable firm_id "Firm ID" ;

. label variable areacode "Area Code up to the Level of Subdistrict";

. label variable industry "4-digits Industrial Sector Code (GB2002)";

. label variable type "Ownership Rights of the Firm";

. label variable opr_hours "Total Annual Operating Hours";

.  label variable product "Total VALUE of Output (in RMB10000)";

. label variable quantity1 "Total QUANTITY of Output 1";

. label variable quantity2 "Total QUANTITY of Output 2";

. label variable quantity3 "Total QUANTITY of Output 3";

. label variable quantity4 "Total QUANTITY of Output 4";

. label variable quantity5 "Total QUANTITY of Output 5";

. label variable wastewater_g ///
>         "Total Amount of Wastewater Generated (in ton)";

. label variable wastewater_e ///
>         "Total Amount of Wastewater Discharged (in ton)";

. label variable water_u ///
>         "Total Water Used = Acquired + Recycled (in ton)";

. label variable water_r ///
>         "Total Water Recycled (in ton)";

. label variable cod_g ///
>         "Total Amount of Chemical Oxygen Demand Generated (in ton)";

. label variable cod_e ///
>         "Total Amount of Chemical Oxygen Demand Discharged (in ton)";

. label variable nh_g ///
>         "Total Amount of Ammonian Generated (in ton)";
```

```
. label variable nh_e ///
>         "Total Amount of Ammonian Discharged (in ton)";

. label variable pet_g ///
>         "Total Amount of Petroleum Generated (in ton)";

. label variable pet_e ///
>         "Total Amount of Petroleum Discharged (in ton)";

.  label variable phe_g ///
>         "Total Amount of Volatile Phenol Generated (in ton)";

. label variable phe_e ///
>         "Total Amount of Volatile Phenol Discharged (in ton)";

. label variable bod_g ///
>         "Total Amount of Biochemical Oxygen Demand Generated (in ton)";

. label variable bod_e ///
>         "Total Amount of Biochemical Oxygen Demand Discharged (in ton)";

. label variable cyn_g ///
>         "Total Amount of Cyanidum Generated (in kg)";

. label variable cyn_e ///
>         "Total Amount of Cyanidum Discharged (in kg)";

. label variable as_g ///
>         "Total Amount of Arsenium Generated (in kg)";

. label variable as_e ///
>         "Total Amount of Arsenium Discharged (in kg)";

. label variable chr_g ///
>         "Total Amount of Chromium Generated (in kg)";

. label variable chr_e ///
>         "Total Amount of Chromium Discharged (in kg)";

. label variable chr6_g ///
>         "Total Amount of Hexavalent Chrome Generated (in kg)";

. label variable chr6_e ///
>         "Total Amount of Hexavalent Chrome Discharged (in kg)";

. label variable dm1_inv ///
>         "Total Investment of Wastewater Disposal Equipment 1 (RMB 10000)";

. label variable dm1_quant ///
>         "Designed Disposal Capacity of Equip 1 (in ton)";
```

```
. label variable dm1_oprcost ///
>         "Total Operating Costs of Equip 1 (RMB 10000)";

. label variable dm1_hours ///
>         "Total Operating Hours of Equip 1";

. label variable dm1_elec ///
>         "Total Amount of Electricity Consumed Equip 1 (10000kw/h)";

. label variable dm1_ef ///
>         "Efficiency of Equip 1 (in %)";

. label variable dm1_code ///
>         "Code for Disposal Method of Equip 1";

. label variable dm2_inv ///
>         "Total Investment of Wastewater Disposal Equipment 2 (RMB 10000)";

. label variable dm2_quant ///
>         "Designed Disposal Capacity of Equip 2 (in ton)";

. label variable dm2_oprcost ///
>         "Total Operating Costs of Equip 2 (RMB 10000)";

. label variable dm2_hours ///
>         "Total Operating Hours of Equip 2";

. label variable dm2_elec ///
>         "Total Amount of Electricity Consumed Equip 2 (10000kw/h)";

. label variable dm2_ef ///
>         "Efficiency of Equip 2 (in %)";

. label variable dm2_code ///
>         "Code for Disposal Method of Equip 2";

. label variable dm3_inv ///
>         "Total Investment of Wastewater Disposal Equipment 3 (RMB 10000)";

. label variable dm3_quant ///
>         "Designed Disposal Capacity of Equip 3 (in ton)";

. label variable dm3_oprcost ///
>         "Total Operating Costs of Equip 3 (RMB 10000)";

. label variable dm3_hours ///
>         "Total Operating Hours of Equip 3";

. label variable dm3_elec ///
```

```
>         "Total Amount of Electricity Consumed Equip 3 (10000kw/h)";

. label variable dm3_ef ///
>         "Efficiency of Equip 3 (in %)";

. label variable dm3_code ///
>         "Code for Disposal Method of Equip 3";

. label variable Census_Type ///
>         "Census Type Code: 1 Key 2 Regular";

. * Aggregate counties to provinces ;
. recode areacode (110000000000/119999999999 = 11)
>         (120000000000/129999999999 = 12) (130000000000/139999999999 = 13)
>         (140000000000/149999999999 = 14) (150000000000/159999999999 = 15)
>         (210000000000/219999999999 = 21) (220000000000/229999999999 = 22)
>         (230000000000/239999999999 = 23) (310000000000/319999999999 = 31)
>         (320000000000/329999999999 = 32) (330000000000/339999999999 = 33)
>         (340000000000/349999999999 = 34) (350000000000/359999999999 = 35)
>         (360000000000/369999999999 = 36) (370000000000/379999999999 = 37)
>         (410000000000/419999999999 = 41) (420000000000/429999999999 = 42)
>         (430000000000/439999999999 = 43) (440000000000/449999999999 = 44)
>         (450000000000/459999999999 = 45) (460000000000/469999999999 = 46)
>         (500000000000/509999999999 = 50) (510000000000/519999999999 = 51)
>         (520000000000/529999999999 = 52) (530000000000/539999999999 = 53)
>         (540000000000/549999999999 = 54) (610000000000/619999999999 = 61)
>         (620000000000/629999999999 = 62) (630000000000/639999999999 = 63)
>         (640000000000/649999999999 = 64) (650000000000/659999999999 = 65),
>         generate(province) ;
(921004 differences between areacode and province)

. * Aggregate sectors ;
. recode industry (500/599 = 5)
>         (600/699 = 6) (700/799 = 7) (800/899 = 8) (900/999 = 9)
>         (1000/1099 = 10) (1100/1199 = 11) (1300/1399 = 13) (1400/1499 = 14)
>         (1500/1599 = 15) (1600/1699 = 16) (1700/1799 = 17) (1800/1899 = 18)
>         (1900/1999 = 19) (2000/2099 = 20) (2100/2199 = 21) (2200/2299 = 22)
>         (2300/2399 = 23) (2400/2499 = 24) (2500/2599 = 25) (2600/2699 = 26)
>         (2700/2799 = 27) (2800/2899 = 28) (2900/2999 = 29) (3000/3099 = 30)
>         (3100/3199 = 31) (3200/3299 = 32) (3300/3399 = 33) (3400/3499 = 34)
>         (3500/3599 = 35) (3600/3699 = 36) (3700/3799 = 37) (3900/3999 = 39)
>         (4000/4099 = 40) (4100/4199 = 41) (4200/4299 = 42) (4300/4399 = 43)
>         (4400/4499 = 44) (4500/4599 = 45) (4600/4699 = 46),
>         generate(industry_a) ;
(921004 differences between industry and industry_a)

. save allfirms.dta, replace ;
file allfirms.dta saved

. log close ;
```

```
      name:  <unnamed>
  log type:  text
 closed on:  18 Sep 2019, 15:54:04
-------------------------------------------------------------------------------
-------------------------------------------------------------------------------
----------
```