```
> # ------------------------------------------------------------------------
> #                       Program Description
> # -------------------- .... [TRUNCATED]

> # Load in packages
> library(foreign)

> library(data.table)

> # ------------------------------------------------------------------------
> #                           Key Firms
> # ------------------------- .... [TRUNCATED]

> # For faster processing, save data as data.table
> KEYFIRM <- data.table(KEYFIRM, key=c("firm_id"))

> comment(KEYFIRM$firm_id) <- "Firm ID"

> comment(KEYFIRM$areacode) <- "Area Code up to the Level of Subdistrict"

> comment(KEYFIRM$industry) <- "4-digits Industrial Sector Code (GB2002)"

> comment(KEYFIRM$type) <- "Ownership Rights of the Firm"

> comment(KEYFIRM$opr_hours) <- "Total Annual Operating Hours"

> comment(KEYFIRM$product) <- "Total VALUE of Output (in RMB10000)"

> comment(KEYFIRM$quantity1) <- "Total QUANTITY of Output 1"

> comment(KEYFIRM$quantity2) <- "Total QUANTITY of Output 2"

> comment(KEYFIRM$quantity3) <- "Total QUANTITY of Output 3"

> comment(KEYFIRM$quantity4) <- "Total QUANTITY of Output 4"

> comment(KEYFIRM$quantity5) <- "Total QUANTITY of Output 5"

> comment(KEYFIRM$wastewater_g) <-
+   "Total Amount of Wastewater Generated (in ton)"

> comment(KEYFIRM$wastewater_e) <-
+   "Total Amount of Wastewater Discharged (in ton)"

> comment(KEYFIRM$water_u) <-
+   "Total Water Used = Acquired + Recycled (in ton)"

> comment(KEYFIRM$water_r) <- "Total Water Recycled (in ton)"
```

```
> comment(KEYFIRM$cod_g) <-
+   "Total Amount of Chemical Oxygen Demand Generated (in ton)"

> comment(KEYFIRM$cod_e) <-
+   "Total Amount of Chemical Oxygen Demand Discharged (in ton)"

> comment(KEYFIRM$nh_g) <- "Total Amount of Ammonian Generated (in ton)"

> comment(KEYFIRM$nh_e) <- "Total Amount of Ammonian Discharged (in ton)"

> comment(KEYFIRM$pet_g) <- "Total Amount of Petroleum Generated (in ton)"

> comment(KEYFIRM$pet_e) <- "Total Amount of Petroleum Discharged (in ton)"

> comment(KEYFIRM$phe_g) <-
+   "Total Amount of Volatile Phenol Generated (in ton)"

> comment(KEYFIRM$phe_e) <-
+   "Total Amount of Volatile Phenol Discharged (in ton)"

> comment(KEYFIRM$bod_g) <-
+   "Total Amount of Biochemical Oxygen Demand Generated (in ton)"

> comment(KEYFIRM$bod_e) <-
+   "Total Amount of Biochemical Oxygen Demand Discharged (in ton)"

> comment(KEYFIRM$cyn_g) <- "Total Amount of Cyanidum Generated (in kg)"

> comment(KEYFIRM$cyn_e) <- "Total Amount of Cyanidum Discharged (in kg)"

> comment(KEYFIRM$as_g) <- "Total Amount of Arsenium Generated (in kg)"

> comment(KEYFIRM$as_e) <- "Total Amount of Arsenium Discharged (in kg)"

> comment(KEYFIRM$chr_g) <- "Total Amount of Chromium Generated (in kg)"

> comment(KEYFIRM$chr_e) <- "Total Amount of Chromium Discharged (in kg)"

> comment(KEYFIRM$chr6_g) <-
+   "Total Amount of Hexavalent Chrome Generated (in kg)"

> comment(KEYFIRM$chr6_e) <-
+   "Total Amount of Hexavalent Chrome Discharged (in kg)"

> comment(KEYFIRM$dm1_inv) <-
+   "Total Investment of Wastewater Disposal Equipment 1 (RMB 10000)"

> comment(KEYFIRM$dm1_quant) <-
+   "Designed Disposal Capacity of Equip 1 (in ton)"
```

```
> comment(KEYFIRM$dm1_oprcost) <-
+   "Total Operating Costs of Equip 1 (RMB 10000)"

> comment(KEYFIRM$dm1_hours) <-
+   "Total Operating Hours of Equip 1"

> comment(KEYFIRM$dm1_elec) <-
+   "Total Amount of Electricity Consumed Equip 1 (10000kw/h)"

> comment(KEYFIRM$dm1_ef) <- "Efficiency of Equip 1 (in %)"

> comment(KEYFIRM$dm1_code) <- "Code for Disposal Method of Equip 1"

> comment(KEYFIRM$dm2_inv) <-
+   "Total Investment of Wastewater Disposal Equipment 2 (RMB 10000)"

> comment(KEYFIRM$dm2_quant) <-
+   "Designed Disposal Capacity of Equip 2 (in ton)"

> comment(KEYFIRM$dm2_oprcost) <-
+   "Total Operating Costs of Equip 2 (RMB 10000)"

> comment(KEYFIRM$dm2_hours) <- "Total Operating Hours of Equip 2"

> comment(KEYFIRM$dm2_elec) <-
+   "Total Amount of Electricity Consumed Equip 2 (10000kw/h)"

> comment(KEYFIRM$dm2_ef) <- "Efficiency of Equip 2 (in %)"

> comment(KEYFIRM$dm2_code) <- "Code for Disposal Method of Equip 2"

> comment(KEYFIRM$dm3_inv) <-
+   "Total Investment of Wastewater Disposal Equipment 3 (RMB 10000)"

> comment(KEYFIRM$dm3_quant) <-
+   "Designed Disposal Capacity of Equip 3 (in ton)"

> comment(KEYFIRM$dm3_oprcost) <-
+   "Total Operating Costs of Equip 3 (RMB 10000)"

> comment(KEYFIRM$dm3_hours) <- "Total Operating Hours of Equip 3"

> comment(KEYFIRM$dm3_elec) <-
+   "Total Amount of Electricity Consumed Equip 3 (10000kw/h)"

> comment(KEYFIRM$dm3_ef) <- "Efficiency of Equip 3 (in %)"

> comment(KEYFIRM$dm3_code) <- "Code for Disposal Method of Equip 3"

> comment(KEYFIRM$Census_Type) <- "Census Type Code: 1 Key 2 Regular"
```

```
> # Save in R's internal binary form
> save(KEYFIRM, file = "./Data/KEYFIRM_R.RData")

> rm(KEYFIRM)

> # ---------------------------------------------------------------------------
> #                            Regular Firms
> # ------------------------ .... [TRUNCATED]

> # For faster processing, save data as data.table
> REGFIRM <- data.table(REGFIRM, key=c("firm_id"))

> comment(REGFIRM$firm_id) <- "Firm ID"

> comment(REGFIRM$areacode) <- "Area Code up to the Level of Subdistrict"

> comment(REGFIRM$industry) <- "4-digits Industrial Sector Code (GB2002)"

> comment(REGFIRM$type) <- "Ownership Rights of the Firm"

> comment(REGFIRM$opr_hours) <- "Total Annual Operating Hours"

> comment(REGFIRM$product) <- "Total VALUE of Output (in RMB10000)"

> comment(REGFIRM$quantity1) <- "Total QUANTITY of Output 1"

> comment(REGFIRM$quantity2) <- "Total QUANTITY of Output 2"

> comment(REGFIRM$quantity3) <- "Total QUANTITY of Output 3"

> comment(REGFIRM$quantity4) <- "Total QUANTITY of Output 4"

> comment(REGFIRM$quantity5) <- "Total QUANTITY of Output 5"

> comment(REGFIRM$wastewater_g) <-
+   "Total Amount of Wastewater Generated (in ton)"

> comment(REGFIRM$wastewater_e) <-
+   "Total Amount of Wastewater Discharged (in ton)"

> comment(REGFIRM$water_u) <-
+   "Total Water Used = Acquired + Recycled (in ton)"

> comment(REGFIRM$water_r) <- "Total Water Recycled (in ton)"

> comment(REGFIRM$cod_g) <-
+   "Total Amount of Chemical Oxygen Demand Generated (in ton)"

> comment(REGFIRM$cod_e) <-
```

```
+   "Total Amount of Chemical Oxygen Demand Discharged (in ton)"

> comment(REGFIRM$nh_g) <- "Total Amount of Ammonian Generated (in ton)"

> comment(REGFIRM$nh_e) <- "Total Amount of Ammonian Discharged (in ton)"

> comment(REGFIRM$pet_g) <- "Total Amount of Petroleum Generated (in ton)"

> comment(REGFIRM$pet_e) <- "Total Amount of Petroleum Discharged (in ton)"

> comment(REGFIRM$phe_g) <-
+   "Total Amount of Volatile Phenol Generated (in ton)"

> comment(REGFIRM$phe_e) <-
+   "Total Amount of Volatile Phenol Discharged (in ton)"

> comment(REGFIRM$bod_g) <-
+   "Total Amount of Biochemical Oxygen Demand Generated (in ton)"

> comment(REGFIRM$bod_e) <-
+   "Total Amount of Biochemical Oxygen Demand Discharged (in ton)"

> comment(REGFIRM$cyn_g) <- "Total Amount of Cyanidum Generated (in kg)"

> comment(REGFIRM$cyn_e) <- "Total Amount of Cyanidum Discharged (in kg)"

> comment(REGFIRM$as_g) <- "Total Amount of Arsenium Generated (in kg)"

> comment(REGFIRM$as_e) <- "Total Amount of Arsenium Discharged (in kg)"

> comment(REGFIRM$dm1_inv) <-
+   "Total Investment of Wastewater Disposal Equipment 1 (RMB 10000)"

> comment(REGFIRM$dm1_quant) <-
+   "Designed Disposal Capacity of Equip 1 (in ton)"

> comment(REGFIRM$dm1_oprcost) <-
+   "Total Operating Costs of Equip 1 (RMB 10000)"

> comment(REGFIRM$dm1_elec) <-
+   "Total Amount of Electricity Consumed Equip 1 (10000kw/h)"

> comment(REGFIRM$dm1_code) <- "Code for Disposal Method of Equip 1"

> comment(REGFIRM$Census_Type) <- "Census Type Code: 1 Key 2 Regular"

> # Save in R internal binary form
> save(REGFIRM, file = "./Data/REGFIRM_R.RData")

> rm(REGFIRM)
```

```
> # ------------------------------------------------------------------------
> #                            All Firms
> # --------------------------- .... [TRUNCATED]

> # For faster processing, save data as data.table
> ALLFIRM <- data.table(ALLFIRM, key=c("firm_id"))

> comment(ALLFIRM$firm_id) <- "Firm ID"

> comment(ALLFIRM$areacode) <- "Area Code up to the Level of Subdistrict"

> comment(ALLFIRM$industry) <- "4-digits Industrial Sector Code (GB2002)"

> comment(ALLFIRM$type) <- "Ownership Rights of the Firm"

> comment(ALLFIRM$opr_hours) <- "Total Annual Operating Hours"

> comment(ALLFIRM$product) <- "Total VALUE of Output (in RMB10000)"

> comment(ALLFIRM$quantity1) <- "Total QUANTITY of Output 1"

> comment(ALLFIRM$quantity2) <- "Total QUANTITY of Output 2"

> comment(ALLFIRM$quantity3) <- "Total QUANTITY of Output 3"

> comment(ALLFIRM$quantity4) <- "Total QUANTITY of Output 4"

> comment(ALLFIRM$quantity5) <- "Total QUANTITY of Output 5"

> comment(ALLFIRM$wastewater_g) <-
+   "Total Amount of Wastewater Generated (in ton)"

> comment(ALLFIRM$wastewater_e) <-
+   "Total Amount of Wastewater Discharged (in ton)"

> comment(ALLFIRM$water_u) <-
+   "Total Water Used = Acquired + Recycled (in ton)"

> comment(ALLFIRM$water_r) <- "Total Water Recycled (in ton)"

> comment(ALLFIRM$cod_g) <-
+   "Total Amount of Chemical Oxygen Demand Generated (in ton)"

> comment(ALLFIRM$cod_e) <-
+   "Total Amount of Chemical Oxygen Demand Discharged (in ton)"

> comment(ALLFIRM$nh_g) <- "Total Amount of Ammonian Generated (in ton)"

> comment(ALLFIRM$nh_e) <- "Total Amount of Ammonian Discharged (in ton)"
```

```
> comment(ALLFIRM$pet_g) <- "Total Amount of Petroleum Generated (in ton)"

> comment(ALLFIRM$pet_e) <- "Total Amount of Petroleum Discharged (in ton)"

> comment(ALLFIRM$phe_g) <-
+   "Total Amount of Volatile Phenol Generated (in ton)"

> comment(ALLFIRM$phe_e) <-
+   "Total Amount of Volatile Phenol Discharged (in ton)"

> comment(ALLFIRM$bod_g) <-
+   "Total Amount of Biochemical Oxygen Demand Generated (in ton)"

> comment(ALLFIRM$bod_e) <-
+   "Total Amount of Biochemical Oxygen Demand Discharged (in ton)"

> comment(ALLFIRM$cyn_g) <- "Total Amount of Cyanidum Generated (in kg)"

> comment(ALLFIRM$cyn_e) <- "Total Amount of Cyanidum Discharged (in kg)"

> comment(ALLFIRM$as_g) <- "Total Amount of Arsenium Generated (in kg)"

> comment(ALLFIRM$as_e) <- "Total Amount of Arsenium Discharged (in kg)"

> comment(ALLFIRM$chr_g) <- "Total Amount of Chromium Generated (in kg)"

> comment(ALLFIRM$chr_e) <- "Total Amount of Chromium Discharged (in kg)"

> comment(ALLFIRM$chr6_g) <-
+   "Total Amount of Hexavalent Chrome Generated (in kg)"

> comment(ALLFIRM$chr6_e) <-
+   "Total Amount of Hexavalent Chrome Discharged (in kg)"

> comment(ALLFIRM$dm1_inv) <-
+   "Total Investment of Wastewater Disposal Equipment 1 (RMB 10000)"

> comment(ALLFIRM$dm1_quant) <-
+   "Designed Disposal Capacity of Equip 1 (in ton)"

> comment(ALLFIRM$dm1_oprcost) <-
+   "Total Operating Costs of Equip 1 (RMB 10000)"

> comment(ALLFIRM$dm1_hours) <- "Total Operating Hours of Equip 1"

> comment(ALLFIRM$dm1_elec) <-
+   "Total Amount of Electricity Consumed Equip 1 (10000kw/h)"

> comment(ALLFIRM$dm1_ef) <- "Efficiency of Equip 1 (in %)"
```

```
> comment(ALLFIRM$dm1_code) <- "Code for Disposal Method of Equip 1"

> comment(ALLFIRM$dm2_inv) <-
+   "Total Investment of Wastewater Disposal Equipment 2 (RMB 10000)"

> comment(ALLFIRM$dm2_quant) <-
+   "Designed Disposal Capacity of Equip 2 (in ton)"

> comment(ALLFIRM$dm2_oprcost) <-
+   "Total Operating Costs of Equip 2 (RMB 10000)"

> comment(ALLFIRM$dm2_hours) <- "Total Operating Hours of Equip 2"

> comment(ALLFIRM$dm2_elec) <-
+   "Total Amount of Electricity Consumed Equip 2 (10000kw/h)"

> comment(ALLFIRM$dm2_ef) <- "Efficiency of Equip 2 (in %)"

> comment(ALLFIRM$dm2_code) <- "Code for Disposal Method of Equip 2"

> comment(ALLFIRM$dm3_inv) <-
+   "Total Investment of Wastewater Disposal Equipment 3 (RMB 10000)"

> comment(ALLFIRM$dm3_quant) <-
+   "Designed Disposal Capacity of Equip 3 (in ton)"

> comment(ALLFIRM$dm3_oprcost) <-
+   "Total Operating Costs of Equip 3 (RMB 10000)"

> comment(ALLFIRM$dm3_hours) <-
+   "Total Operating Hours of Equip 3"

> comment(ALLFIRM$dm3_elec) <-
+   "Total Amount of Electricity Consumed Equip 3 (10000kw/h)"

> comment(ALLFIRM$dm3_ef) <- "Efficiency of Equip 3 (in %)"

> comment(ALLFIRM$dm3_code) <- "Code for Disposal Method of Equip 3"

> comment(ALLFIRM$Census_Type) <- "Census Type Code: 1 Key 2 Regular"

> # Save in R's internal binary form
> save(ALLFIRM, file = "./Data/ALLFIRM_R.RData")

> rm(ALLFIRM)
```