Data Mining Project

Khalid

Data Mining Project

1. Open file in R

```
data <- read.csv("C:/Users/Khalid Laptop/Desktop/Data Mining Project/visa_number_in_japan.csv")
```

2. Look into the content of the data

```
str(data)
```

```
## 'data.frame':    1933 obs. of  48 variables:
##  $ Year                                  : int  2017 2017 2017 2017 2017 2017 2017 2017 2017 2017 ...
##  $ Regional.code                         : int  40 80 120 200 240 280 310 320 360 400 ...
##  $ Country                               : chr  "Afghanistan" "Albania" "Algeria" "Andra" ...
##  $ Number.of.issued                      : int  1005 252 899 5 111 41 619 450 3887 875 ...
##  $ Number.of.issued_numerical            : int  46 7 19 0 12 8 8 0 281 0 ...
##  $ Travel.certificate                    : int  0 0 0 0 75 0 0 0 0 0 ...
##  $ Diplomacy                             : int  14 0 8 0 9 1 31 0 54 0 ...
##  $ Public.use                            : int  112 2 11 0 7 0 20 1 220 0 ...
##  $ Passing                               : int  0 1 3 0 0 0 0 0 0 0 ...
##  $ Short..term.stay                      : int  224 200 813 0 67 28 505 0 5 0 ...
##  $ Staying.in.medical.care               : int  0 0 0 0 0 0 0 0 0 0 ...
##  $ Advanced.profession                   : int  0 0 0 0 0 0 1 0 8 0 ...
##  $ Employment                            : int  38 8 25 3 0 0 14 209 1501 684 ...
##  $ Employment_Professor                  : int  0 1 2 0 0 0 2 3 86 21 ...
```

```
##  $ Employment_Art                            : int  0 0 0 0 0 0 0 44 6 0 ...
##  $ Employment_religion                       : int  0 0 0 0 0 0 0 2 22 0 ...
##  $ Employment_Report                         : int  0 0 0 0 0 0 1 0 3 0 ...
##  $ Employment_Management...Management       : int  0 0 0 0 0 0 0 0 33 1 ...
##  $ Employment_Law.accounting                 : int  0 0 0 0 0 0 0 0 0 0 ...
##  $ Employment_Medical.care                   : int  0 0 0 0 0 0 0 0 0 0 ...
##  $ Employment_Research                       : int  0 0 0 0 0 0 0 0 9 1 ...
##  $ Employment_Education                      : int  0 0 0 0 0 0 0 0 210 0 ...
##  $ Employment_Country                        : int  37 2 22 0 0 0 6 6 286 15 ...
##  $ Employment_Technology                     : int  NA NA NA NA NA NA NA NA NA NA ...
##  $ Employment_Humanities.International       : int  NA NA NA NA NA NA NA NA NA NA ...
##  $ Employment_C.corporate.rotation           : int  0 1 0 0 0 0 1 5 40 11 ...
##  $ Employment_show                           : int  0 2 1 2 0 0 4 129 692 623 ...
##  $ Employment_Skills                         : int  1 2 0 1 0 0 0 20 114 12 ...
##  $ Employment_Steng..new.                    : int  0 0 0 0 0 0 0 0 0 0 ...
##  $ Employment_Nursing.care                   : int  0 0 0 0 0 0 0 0 0 0 ...
##  $ Ordinary                                  : int  604 41 37 2 28 12 46 147 758 136 ...
##  $ General_Cultural.activity                 : int  1 4 2 0 0 0 1 11 21 7 ...
##  $ General_Studying.abroad                   : int  66 9 14 2 3 0 22 33 586 119 ...
##  $ General_Arademy                           : int  NA NA NA NA NA NA NA NA NA NA ...
##  $ General_Training                          : int  248 27 13 0 25 12 22 85 21 0 ...
##  $ General_Family.stay                       : int  289 1 8 0 0 0 1 13 130 10 ...
##  $ General_Stifying.1.I                      : int  0 0 0 0 0 0 0 5 0 0 ...
##  $ General_Steng.1                           : int  0 0 0 0 0 0 0 0 0 0 ...
##  $ identification                            : int  13 0 2 0 0 0 2 93 1341 55 ...
##  $ Specific_housework.employee               : int  0 0 0 0 0 0 0 0 0 0 ...
##  $ Specified_short.term                      : int  NA NA NA NA NA NA NA NA NA NA ...
##  $ Specific_profit.representative.staff      : int  0 0 0 0 0 0 0 0 0 0 ...
##  $ Specific_working.holiday                  : int  0 0 0 0 0 0 0 45 1182 49 ...
##  $ Specified_Amasport                        : int  0 0 0 0 0 0 0 0 21 0 ...
##  $ Specific_Japanese.spouse..etc.            : int  2 0 1 0 0 0 1 19 114 5 ...
##  $ Specific_Permanent.resident.s.spouse..etc.: int  3 0 1 0 0 0 1 2 4 0 ...
##  $ Specified_Distingant                      : int  8 0 0 0 0 0 0 25 2 0 ...
##  $ Specific_Others                           : int  0 0 0 0 0 0 0 2 18 1 ...
```

3. Data Cleaning

```r
years <- unique(data$Year)
countries <- unique(data$Country)

for (year in years) {
  assign(paste0("data_", year), subset(data, Year == year))
}

for (country in countries) {
  country_name <- gsub("[^a-zA-Z0-9]", "_", country)
  assign(paste0("data_", country_name), subset(data, Country == country))
}

ls(pattern = "data_")
```

3.1 Separate the data into their respective year and country

```
##   [1] "data_2006"
##   [2] "data_2007"
##   [3] "data_2008"
##   [4] "data_2009"
##   [5] "data_2010"
##   [6] "data_2011"
##   [7] "data_2012"
##   [8] "data_2013"
##   [9] "data_2014"
##  [10] "data_2015"
##  [11] "data_2016"
##  [12] "data_2017"
##  [13] "data_Afghanistan"
##  [14] "data_Albania"
##  [15] "data_Algeria"
##  [16] "data_Andra"
##  [17] "data_Angola"
##  [18] "data_Antigua_Berbuda"
##  [19] "data_Argentina"
##  [20] "data_Armenia"
##  [21] "data_Australia"
```

```
##  [22] "data_Austria"
##  [23] "data_Azerbaijan"
##  [24] "data_Bahamas"
##  [25] "data_Bangladesh"
##  [26] "data_Barbados"
##  [27] "data_Barrane"
##  [28] "data_Belarus"
##  [29] "data_Belgium"
##  [30] "data_Belize"
##  [31] "data_Benan"
##  [32] "data_Bhutan"
##  [33] "data_Bolivia"
##  [34] "data_Bosnia_Herzegovna"
##  [35] "data_Botwana"
##  [36] "data_Brazil"
##  [37] "data_British"
##  [38] "data_British__other_"
##  [39] "data_British_overseas"
##  [40] "data_British_protection"
##  [41] "data_British_territories"
##  [42] "data_Brunei"
##  [43] "data_Brush"
##  [44] "data_Bulgaria"
##  [45] "data_Burkina_Faso"
##  [46] "data_Cambodia"
##  [47] "data_Cameroon"
##  [48] "data_Canada"
##  [49] "data_Carbo_verde"
##  [50] "data_Central_Africa"
##  [51] "data_Chad"
##  [52] "data_Chile"
##  [53] "data_China"
##  [54] "data_China__Hong_Kong_"
##  [55] "data_Columbia"
##  [56] "data_Congo"
##  [57] "data_Congo_Democratic_Republic"
##  [58] "data_Costa_Rica"
##  [59] "data_Court_Jiboire"
##  [60] "data_Croatia"
```

```
##  [61] "data_Cuba"
##  [62] "data_Cyprus"
##  [63] "data_Czech_Republic"
##  [64] "data_Denmark"
##  [65] "data_Dominica"
##  [66] "data_dominican_republic"
##  [67] "data_East_Timor_Democratic_Republic"
##  [68] "data_Ecuador"
##  [69] "data_Egypt"
##  [70] "data_El_Salvador"
##  [71] "data_Equatorial_guinea"
##  [72] "data_Eritria"
##  [73] "data_Estonia"
##  [74] "data_Ethiopia"
##  [75] "data_Fiji"
##  [76] "data_Finland"
##  [77] "data_Former_Macedonia_Yugoslavia_Republic"
##  [78] "data_France"
##  [79] "data_Gabon"
##  [80] "data_Gaiana"
##  [81] "data_Gambia"
##  [82] "data_Georgia"
##  [83] "data_Germany"
##  [84] "data_Ghana"
##  [85] "data_Greece"
##  [86] "data_Grenada"
##  [87] "data_Guatemala"
##  [88] "data_Guinea"
##  [89] "data_Guinea_Visau"
##  [90] "data_Haiti"
##  [91] "data_Honduras"
##  [92] "data_Hong_Kong"
##  [93] "data_Hong_Kong__UK_"
##  [94] "data_Hong_Kong_DI"
##  [95] "data_Hong_Kong_SAR"
##  [96] "data_Hungary"
##  [97] "data_Iceland"
##  [98] "data_India"
##  [99] "data_Indonesia"
```

```
## [100] "data_Iran"
## [101] "data_Iraq"
## [102] "data_Ireland"
## [103] "data_Israel"
## [104] "data_Italy"
## [105] "data_Jamaica"
## [106] "data_Jibuti"
## [107] "data_Jordan"
## [108] "data_Kazakhstan"
## [109] "data_Kenya"
## [110] "data_Kiribass"
## [111] "data_Komoro"
## [112] "data_Kosovo"
## [113] "data_Kuwait"
## [114] "data_Kyrgyzstan"
## [115] "data_Laos"
## [116] "data_Latvia"
## [117] "data_Lebanon"
## [118] "data_Lesot"
## [119] "data_Liberia"
## [120] "data_Libya"
## [121] "data_Liechtenstein"
## [122] "data_Lithuania"
## [123] "data_Luxembourg"
## [124] "data_Macau"
## [125] "data_Macau_travel_certificate"
## [126] "data_Madagascar"
## [127] "data_Malawi"
## [128] "data_Malaysia"
## [129] "data_Maldives"
## [130] "data_Malta"
## [131] "data_Mari"
## [132] "data_Marshall"
## [133] "data_Mauritania"
## [134] "data_Mauritius"
## [135] "data_Mexico"
## [136] "data_Micronesia"
## [137] "data_Moldoba"
## [138] "data_Monaco"
```

```
## [139] "data_Mongolia"
## [140] "data_Montenegro"
## [141] "data_Morocco"
## [142] "data_Mozambique"
## [143] "data_Myanmar"
## [144] "data_Namibia"
## [145] "data_Naure"
## [146] "data_Nepal"
## [147] "data_Netherlands"
## [148] "data_new_zealand"
## [149] "data_Nicaragua"
## [150] "data_Nigail"
## [151] "data_Nigeria"
## [152] "data_north_korea"
## [153] "data_Norway"
## [154] "data_Oman"
## [155] "data_others"
## [156] "data_Pakistan"
## [157] "data_Palao"
## [158] "data_Palestine"
## [159] "data_Panama"
## [160] "data_Papua_New_Guinea"
## [161] "data_Paraguay"
## [162] "data_Peru"
## [163] "data_Philippines"
## [164] "data_Poland"
## [165] "data_Portugal"
## [166] "data_Qatar"
## [167] "data_Republic_of_South_Africa"
## [168] "data_Republic_of_Trinidad_and_Tobago"
## [169] "data_Romania"
## [170] "data_Russia"
## [171] "data_Rwanda"
## [172] "data_Saechel"
## [173] "data_Samoa_Independent_State"
## [174] "data_Santa_Principa"
## [175] "data_Saudi_Arabia"
## [176] "data_Senegal"
## [177] "data_Serbia"
```

```
## [178] "data_Sierra_Leone"
## [179] "data_Singapore"
## [180] "data_Slinum"
## [181] "data_Slovakia"
## [182] "data_Slovenia"
## [183] "data_Solomon"
## [184] "data_Somalia"
## [185] "data_South_Korea"
## [186] "data_South_Sudan"
## [187] "data_Spain"
## [188] "data_Sri_Lanka"
## [189] "data_St__Christopher_Navis"
## [190] "data_St__Lucia"
## [191] "data_St__Vincent"
## [192] "data_Sudan"
## [193] "data_Sun_Marino"
## [194] "data_Swaji_Land"
## [195] "data_Sweden"
## [196] "data_Switzerland"
## [197] "data_Syria"
## [198] "data_Taiwan"
## [199] "data_Tajikistan"
## [200] "data_Tanzania"
## [201] "data_Thailand"
## [202] "data_Togo"
## [203] "data_Tonga"
## [204] "data_Torque_menistan"
## [205] "data_Tsubaru"
## [206] "data_Tunisia"
## [207] "data_Turkey"
## [208] "data_Uganda"
## [209] "data_UK_s_directly_under_the_UK"
## [210] "data_Ukraine"
## [211] "data_United_Arab_Emirates"
## [212] "data_united_states_of_america"
## [213] "data_Uruguay"
## [214] "data_Uzbekistan"
## [215] "data_Vanuatsu"
## [216] "data_Vatican"
```

```
## [217] "data_Venezuela"
## [218] "data_Vietnam"
## [219] "data_Yemen"
## [220] "data_Yugoslavia"
## [221] "data_Yugoslavia_Socialist_Federal_Republic"
## [222] "data_Zaire"
## [223] "data_Zambia"
## [224] "data_Zimbabwe"
```

The data is divided into their respective object

```
str(data_2006)
```

To check the object

```
## 'data.frame':    31 obs. of  48 variables:
##  $ Year                                  : int  2006 2006 2006 2006 2006 2006 2006 2006 2006 2006 ...
##  $ Regional.code                         : int  40 360 500 760 1040 1240 1440 1560 1580 2500 ...
##  $ Country                               : chr  "Afghanistan" "Australia" "Bangladesh" "Brazil" ...
##  $ Number.of.issued                      : int  2547 3878 4434 41927 3339 3307 8185 511606 4532 3683 ...
##  $ Number.of.issued_numerical            : int  69 85 93 881 193 639 302 8786 423 210 ...
##  $ Travel.certificate                    : int  NA NA NA NA NA NA NA NA NA NA ...
##  $ Diplomacy                             : int  NA NA NA NA NA NA NA NA NA NA ...
##  $ Public.use                            : int  NA NA NA NA NA NA NA NA NA NA ...
##  $ Passing                               : int  0 0 39 247 26 0 135 971 9 0 ...
##  $ Short..term.stay                      : int  2252 8 2473 13668 2193 10 6585 393857 505 17 ...
##  $ Staying.in.medical.care               : int  NA NA NA NA NA NA NA NA NA NA ...
##  $ Advanced.profession                   : int  NA NA NA NA NA NA NA NA NA NA ...
##  $ Employment                            : int  11 2015 252 958 38 1959 186 13988 775 1878 ...
##  $ Employment_Professor                  : int  NA NA NA NA NA NA NA NA NA NA ...
##  $ Employment_Art                        : int  NA NA NA NA NA NA NA NA NA NA ...
##  $ Employment_religion                   : int  NA NA NA NA NA NA NA NA NA NA ...
##  $ Employment_Report                     : int  NA NA NA NA NA NA NA NA NA NA ...
##  $ Employment_Management...Management    : int  NA NA NA NA NA NA NA NA NA NA ...
##  $ Employment_Law.accounting             : int  NA NA NA NA NA NA NA NA NA NA ...
```

```
##  $ Employment_Medical.care                   : int  NA NA NA NA NA NA NA NA NA NA ...
##  $ Employment_Research                       : int  NA NA NA NA NA NA NA NA NA NA ...
##  $ Employment_Education                      : int  NA NA NA NA NA NA NA NA NA NA ...
##  $ Employment_Country                        : int  NA NA NA NA NA NA NA NA NA NA ...
##  $ Employment_Technology                     : int  NA NA NA NA NA NA NA NA NA NA ...
##  $ Employment_Humanities.International       : int  3 727 30 24 14 740 67 615 117 195 ...
##  $ Employment_C.corporate.rotation           : int  NA NA NA NA NA NA NA NA NA NA ...
##  $ Employment_show                           : int  0 739 0 786 1 612 2 5066 293 1198 ...
##  $ Employment_Skills                         : int  NA NA NA NA NA NA NA NA NA NA ...
##  $ Employment_Steng..new.                    : int  NA NA NA NA NA NA NA NA NA NA ...
##  $ Employment_Nursing.care                   : int  NA NA NA NA NA NA NA NA NA NA ...
##  $ Ordinary                                  : int  190 732 1354 631 812 503 1030 88640 2954 1046 ...
##  $ General_Cultural.activity                 : int  NA NA NA NA NA NA NA NA NA NA ...
##  $ General_Studying.abroad                   : int  NA NA NA NA NA NA NA NA NA NA ...
##  $ General_Aradeny                           : int  1 185 352 28 124 97 260 9488 937 61 ...
##  $ General_Training                          : int  137 10 239 263 403 7 494 62739 128 19 ...
##  $ General_Family.stay                       : int  NA NA NA NA NA NA NA NA NA NA ...
##  $ General_Stifying.1.I                      : int  NA NA NA NA NA NA NA NA NA NA ...
##  $ General_Steng.1                           : int  NA NA NA NA NA NA NA NA NA NA ...
##  $ identification                            : int  1 1080 65 26209 59 777 55 9727 289 653 ...
##  $ Specific_housework.employee               : int  NA NA NA NA NA NA NA NA NA NA ...
##  $ Specified_short.term                      : int  NA NA NA NA NA NA NA NA NA NA ...
##  $ Specific_profit.representative.staff      : int  NA NA NA NA NA NA NA NA NA NA ...
##  $ Specific_working.holiday                  : int  NA NA NA NA NA NA NA NA NA NA ...
##  $ Specified_Amasport                        : int  NA NA NA NA NA NA NA NA NA NA ...
##  $ Specific_Japanese.spouse..etc.            : int  1 136 22 7130 25 114 34 5388 163 95 ...
##  $ Specific_Permanent.resident.s.spouse..etc.: int  NA NA NA NA NA NA NA NA NA NA ...
##  $ Specified_Distingant                      : int  0 1 26 18943 25 7 9 3440 35 0 ...
##  $ Specific_Others                           : int  NA NA NA NA NA NA NA NA NA NA ...
```

```
head(data_2006)
```

```
##      Year Regional.code     Country Number.of.issued Number.of.issued_numerical
## 1903 2006            40 Afghanistan             2547                          69
## 1904 2006           360   Australia             3878                          85
## 1905 2006           500  Bangladesh             4434                          93
## 1906 2006           760      Brazil            41927                         881
## 1907 2006          1040     Myanmar             3339                         193
## 1908 2006          1240      Canada             3307                         639
```

```
##      Travel.certificate Diplomacy Public.use Passing Short..term.stay
## 1903                 NA        NA         NA       0             2252
## 1904                 NA        NA         NA       0                8
## 1905                 NA        NA         NA      39             2473
## 1906                 NA        NA         NA     247            13668
## 1907                 NA        NA         NA      26             2193
## 1908                 NA        NA         NA       0               10
##      Staying.in.medical.care Advanced.profession Employment
## 1903                      NA                  NA         11
## 1904                      NA                  NA       2015
## 1905                      NA                  NA        252
## 1906                      NA                  NA        958
## 1907                      NA                  NA         38
## 1908                      NA                  NA       1959
##      Employment_Professor Employment_Art Employment_religion Employment_Report
## 1903                   NA             NA                   NA                NA
## 1904                   NA             NA                   NA                NA
## 1905                   NA             NA                   NA                NA
## 1906                   NA             NA                   NA                NA
## 1907                   NA             NA                   NA                NA
## 1908                   NA             NA                   NA                NA
##      Employment_Management...Management Employment_Law.accounting
## 1903                                 NA                        NA
## 1904                                 NA                        NA
## 1905                                 NA                        NA
## 1906                                 NA                        NA
## 1907                                 NA                        NA
## 1908                                 NA                        NA
##      Employment_Medical.care Employment_Research Employment_Education
## 1903                      NA                  NA                   NA
## 1904                      NA                  NA                   NA
## 1905                      NA                  NA                   NA
## 1906                      NA                  NA                   NA
## 1907                      NA                  NA                   NA
## 1908                      NA                  NA                   NA
##      Employment_Country Employment_Technology
## 1903                 NA                    NA
## 1904                 NA                    NA
## 1905                 NA                    NA
```

```
## 1906                NA                   NA
## 1907                NA                   NA
## 1908                NA                   NA
##      Employment_Humanities.International Employment_C.corporate.rotation
## 1903                                   3                              NA
## 1904                                 727                              NA
## 1905                                  30                              NA
## 1906                                  24                              NA
## 1907                                  14                              NA
## 1908                                 740                              NA
##      Employment_show Employment_Skills Employment_Steng..new.
## 1903               0                NA                     NA
## 1904             739                NA                     NA
## 1905               0                NA                     NA
## 1906             786                NA                     NA
## 1907               1                NA                     NA
## 1908             612                NA                     NA
##      Employment_Nursing.care Ordinary General_Cultural.activity
## 1903                      NA      190                        NA
## 1904                      NA      732                        NA
## 1905                      NA     1354                        NA
## 1906                      NA      631                        NA
## 1907                      NA      812                        NA
## 1908                      NA      503                        NA
##      General_Studying.abroad General_Arademy General_Training
## 1903                      NA               1              137
## 1904                      NA             185               10
## 1905                      NA             352              239
## 1906                      NA              28              263
## 1907                      NA             124              403
## 1908                      NA              97                7
##      General_Family.stay General_Stifying.1.I General_Steng.1 identification
## 1903                  NA                   NA              NA              1
## 1904                  NA                   NA              NA           1080
## 1905                  NA                   NA              NA             65
## 1906                  NA                   NA              NA          26209
## 1907                  NA                   NA              NA             59
## 1908                  NA                   NA              NA            777
##      Specific_housework.employee Specified_short.term
```

```
## 1903                          NA                   NA
## 1904                          NA                   NA
## 1905                          NA                   NA
## 1906                          NA                   NA
## 1907                          NA                   NA
## 1908                          NA                   NA
##      Specific_profit.representative.staff Specific_working.holiday
## 1903                                   NA                       NA
## 1904                                   NA                       NA
## 1905                                   NA                       NA
## 1906                                   NA                       NA
## 1907                                   NA                       NA
## 1908                                   NA                       NA
##      Specified_Amasport Specific_Japanese.spouse..etc.
## 1903                 NA                             1
## 1904                 NA                           136
## 1905                 NA                            22
## 1906                 NA                          7130
## 1907                 NA                            25
## 1908                 NA                           114
##      Specific_Permanent.resident.s.spouse..etc. Specified_Distingant
## 1903                                         NA                    0
## 1904                                         NA                    1
## 1905                                         NA                   26
## 1906                                         NA                18943
## 1907                                         NA                   25
## 1908                                         NA                    7
##      Specific_Others
## 1903              NA
## 1904              NA
## 1905              NA
## 1906              NA
## 1907              NA
## 1908              NA
```

```
str(data_Afghanistan)
```

```
## 'data.frame':    10 obs. of  48 variables:
```

```
##  $ Year                                  : int  2017 2016 2015 2014 2013 2012 2011 2010 2009 2006
##  $ Regional.code                         : int  40 40 40 40 40 40 40 40 40 40
##  $ Country                               : chr  "Afghanistan" "Afghanistan" "Afghanistan" "Afghanistan" ...
##  $ Number.of.issued                      : int  1005 1009 1420 1042 1054 1241 949 1209 1010 2547
##  $ Number.of.issued_numerical            : int  46 60 35 29 11 69 21 106 28 69
##  $ Travel.certificate                    : int  0 0 0 0 1 3 0 0 0 NA
##  $ Diplomacy                             : int  14 13 25 18 9 60 12 40 15 NA
##  $ Public.use                            : int  112 122 173 61 83 200 39 147 65 NA
##  $ Passing                               : int  0 0 1 0 0 0 3 0 1 0
##  $ Short..term.stay                      : int  224 233 275 214 365 317 384 588 574 2252
##  $ Staying.in.medical.care               : int  0 0 0 0 1 2 0 NA NA NA
##  $ Advanced.profession                   : int  0 0 0 0 0 0 NA NA NA NA
##  $ Employment                            : int  38 29 146 113 96 173 141 91 65 11
##  $ Employment_Professor                  : int  0 1 0 1 0 0 0 0 1 NA
##  $ Employment_Art                        : int  0 0 0 0 0 0 0 0 0 NA
##  $ Employment_religion                   : int  0 0 0 0 0 0 0 0 0 NA
##  $ Employment_Report                     : int  0 0 0 0 0 0 0 0 0 NA
##  $ Employment_Management...Management    : int  0 4 5 12 1 9 12 18 10 NA
##  $ Employment_Law.accounting             : int  0 0 0 0 0 0 0 0 0 NA
##  $ Employment_Medical.care               : int  0 0 0 0 0 0 0 0 0 NA
##  $ Employment_Research                   : int  0 0 0 0 0 0 0 0 0 NA
##  $ Employment_Education                  : int  0 0 0 0 0 0 0 0 0 NA
##  $ Employment_Country                    : int  37 21 138 NA NA NA NA NA NA NA
##  $ Employment_Technology                 : int  NA NA NA 0 0 1 1 1 0 NA
##  $ Employment_Humanities.International   : int  NA NA NA 92 91 155 120 68 50 3
##  $ Employment_C.corporate.rotation       : int  0 1 1 5 3 7 7 4 4 NA
##  $ Employment_show                       : int  0 1 0 1 1 0 0 0 0 0
##  $ Employment_Skills                     : int  1 1 2 2 0 1 1 0 0 NA
##  $ Employment_Steng..new.                : int  0 NA NA NA NA NA NA NA NA NA
##  $ Employment_Nursing.care               : int  0 NA NA NA NA NA NA NA NA NA
##  $ Ordinary                              : int  604 590 793 623 487 487 364 337 276 190
##  $ General_Cultural.activity             : int  1 0 0 1 1 0 1 1 0 NA
##  $ General_Studying.abroad               : int  66 82 122 163 106 74 65 17 19 NA
##  $ General_Arademy                       : int  NA NA NA 0 0 NA 0 0 1 1
##  $ General_Training                      : int  248 205 237 268 204 282 193 230 193 137
##  $ General_Family.stay                   : int  289 303 434 191 176 131 105 89 63 NA
##  $ General_Stifying.1.I                  : int  0 0 0 0 0 0 0 0 NA NA
##  $ General_Steng.1                       : int  0 0 0 0 0 0 0 0 NA NA
##  $ identification                        : int  13 22 7 13 13 2 6 6 14 1
```

```
##  $ Specific_housework.employee               : int  0 0 0 0 1 0 0 2 0 NA
##  $ Specified_short.term                      : int  NA NA NA NA NA NA NA NA NA NA
##  $ Specific_profit.representative.staff      : int  0 0 0 0 0 0 0 0 0 NA
##  $ Specific_working.holiday                  : int  0 0 0 0 0 0 0 0 0 NA
##  $ Specified_Amasport                        : int  0 0 0 0 0 0 0 0 0 NA
##  $ Specific_Japanese.spouse..etc.            : int  2 0 0 1 3 1 0 1 3 1
##  $ Specific_Permanent.resident.s.spouse..etc.: int  3 4 0 4 2 0 1 1 1 NA
##  $ Specified_Distingant                      : int  8 7 4 8 6 1 1 1 9 0
##  $ Specific_Others                           : int  0 11 3 0 1 0 4 1 1 NA
```

```
head(data_Afghanistan)
```

```
##      Year Regional.code     Country Number.of.issued Number.of.issued_numerical
## 1    2017            40 Afghanistan             1005                          46
## 202  2016            40 Afghanistan             1009                          60
## 404  2015            40 Afghanistan             1420                          35
## 605  2014            40 Afghanistan             1042                          29
## 811  2013            40 Afghanistan             1054                          11
## 1018 2012            40 Afghanistan             1241                          69
##      Travel.certificate Diplomacy Public.use Passing Short..term.stay
## 1                      0        14        112       0              224
## 202                    0        13        122       0              233
## 404                    0        25        173       1              275
## 605                    0        18         61       0              214
## 811                    1         9         83       0              365
## 1018                   3        60        200       0              317
##      Staying.in.medical.care Advanced.profession Employment
## 1                          0                   0         38
## 202                        0                   0         29
## 404                        0                   0        146
## 605                        0                   0        113
## 811                        1                   0         96
## 1018                       2                   0        173
##      Employment_Professor Employment_Art Employment_religion Employment_Report
## 1                       0              0                   0                 0
## 202                     1              0                   0                 0
## 404                     0              0                   0                 0
## 605                     1              0                   0                 0
```

```
## 811                    0             0                   0                0
## 1018                   0             0                   0                0
##      Employment_Management...Management Employment_Law.accounting
## 1                                     0                         0
## 202                                   4                         0
## 404                                   5                         0
## 605                                  12                         0
## 811                                   1                         0
## 1018                                  9                         0
##      Employment_Medical.care Employment_Research Employment_Education
## 1                          0                   0                    0
## 202                        0                   0                    0
## 404                        0                   0                    0
## 605                        0                   0                    0
## 811                        0                   0                    0
## 1018                       0                   0                    0
##      Employment_Country Employment_Technology
## 1                    37                    NA
## 202                  21                    NA
## 404                 138                    NA
## 605                  NA                     0
## 811                  NA                     0
## 1018                 NA                     1
##      Employment_Humanities.International Employment_C.corporate.rotation
## 1                                     NA                                0
## 202                                   NA                                1
## 404                                   NA                                1
## 605                                   92                                5
## 811                                   91                                3
## 1018                                 155                                7
##      Employment_show Employment_Skills Employment_Steng..new.
## 1                  0                 1                     0
## 202                1                 1                    NA
## 404                0                 2                    NA
## 605                1                 2                    NA
## 811                1                 0                    NA
## 1018               0                 1                    NA
##      Employment_Nursing.care Ordinary General_Cultural.activity
## 1                          0      604                         1
```

```
## 202                      NA      590                       0
## 404                      NA      793                       0
## 605                      NA      623                       1
## 811                      NA      487                       1
## 1018                     NA      487                       0
##      General_Studying.abroad General_Arademy General_Training
## 1                         66             NA              248
## 202                       82             NA              205
## 404                      122             NA              237
## 605                      163              0              268
## 811                      106              0              204
## 1018                      74             NA              282
##      General_Family.stay General_Stifying.1.I General_Steng.1 identification
## 1                    289                    0               0             13
## 202                  303                    0               0             22
## 404                  434                    0               0              7
## 605                  191                    0               0             13
## 811                  176                    0               0             13
## 1018                 131                    0               0              2
##      Specific_housework.employee Specified_short.term
## 1                              0                  NA
## 202                            0                  NA
## 404                            0                  NA
## 605                            0                  NA
## 811                            1                  NA
## 1018                           0                  NA
##      Specific_profit.representative.staff Specific_working.holiday
## 1                                       0                        0
## 202                                     0                        0
## 404                                     0                        0
## 605                                     0                        0
## 811                                     0                        0
## 1018                                    0                        0
##      Specified_Amasport Specific_Japanese.spouse..etc.
## 1                     0                             2
## 202                   0                             0
## 404                   0                             0
## 605                   0                             1
## 811                   0                             3
```

```
## 1018                  0                                  1
##      Specific_Permanent.resident.s.spouse..etc. Specified_Distingant
## 1                                             3                    8
## 202                                           4                    7
## 404                                           0                    4
## 605                                           4                    8
## 811                                           2                    6
## 1018                                          0                    1
##      Specific_Others
## 1                  0
## 202               11
## 404                3
## 605                0
## 811                1
## 1018               0
```

4. Data Analysis

4.1 The total number of issued for each of the country through out the years

```
library(ggplot2)
```

Example: Malaysia

```
## Warning: package 'ggplot2' was built under R version 4.4.2
```

```
Malaysia_data <- subset(data, Country == "Malaysia")

Malaysia_data$Year <- as.numeric(Malaysia_data$Year)
Malaysia_data$Number.of.issued <- as.numeric(Malaysia_data$Number.of.issued)

ggplot(Malaysia_data, aes(x = Year, y = Number.of.issued)) +
  geom_line(color = "blue", size = 1) +
  geom_point(color = "red", size = 2) +
  labs(title = "Number of Visas Issued for Malaysia (2006-2017)",
```

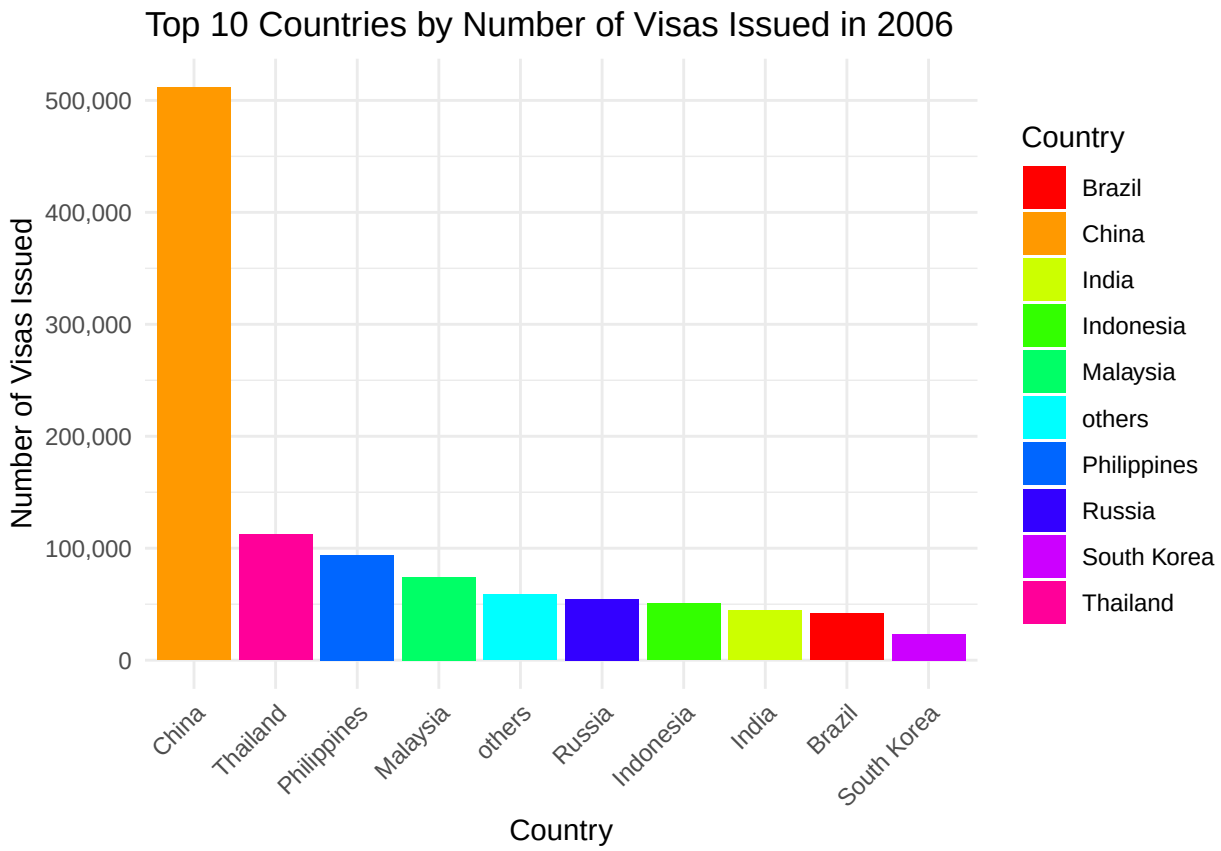
```
      x = "Year",
      y = "Number of Visas Issued") +
  theme_minimal() +
  scale_x_continuous(breaks = seq(2006, 2017, 1)) +
  scale_y_continuous(labels = scales::comma)
```

```
## Warning: Using ‘size‘ aesthetic for lines was deprecated in ggplot2 3.4.0.
## i Please use ‘linewidth‘ instead.
## This warning is displayed once every 8 hours.
## Call ‘lifecycle::last_lifecycle_warnings()‘ to see where this warning was
## generated.
```

4.2 The total number of issued for each of the country in a specific year

```
library(ggplot2)

data_2006 <- subset(data, Year == 2006)

data_2006$Number.of.issued <- as.numeric(data_2006$Number.of.issued)
```

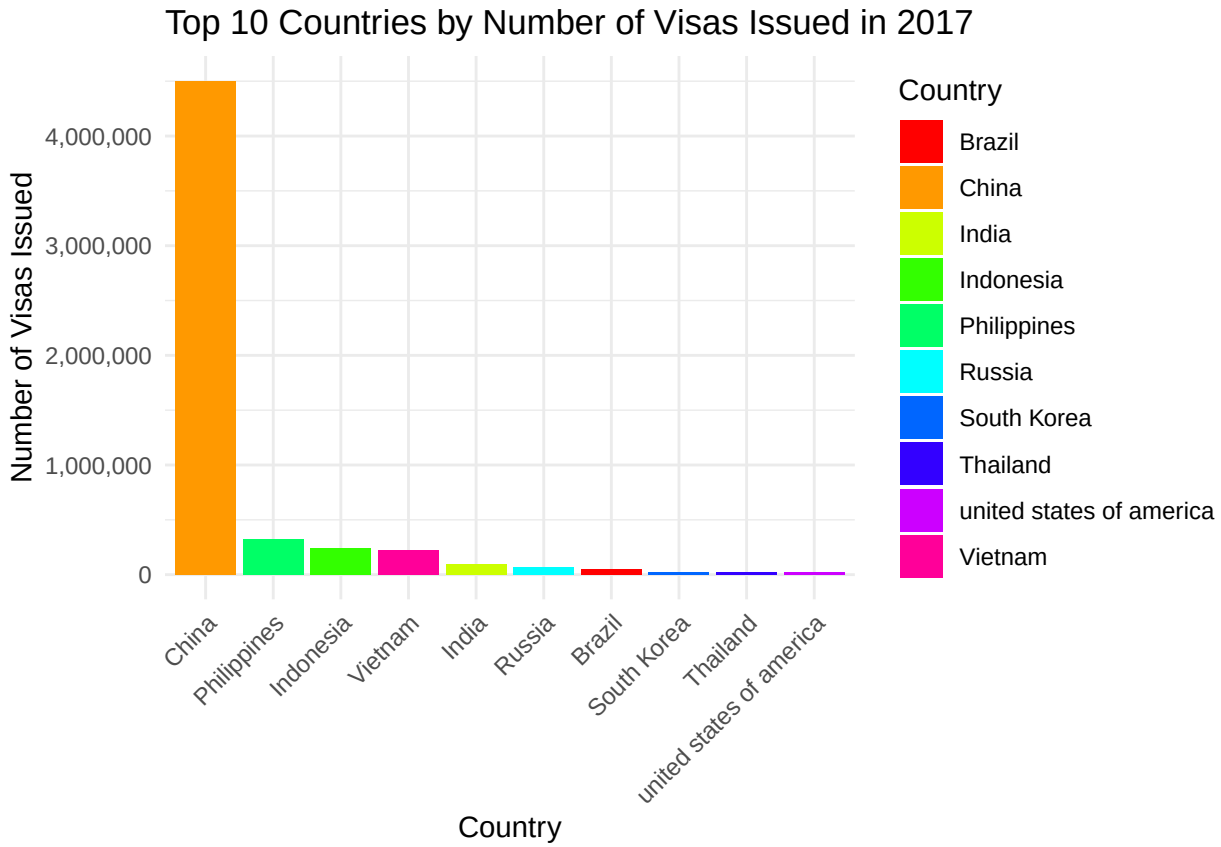
```r
top_10_countries <- data_2006[order(-data_2006$Number.of.issued), ][1:10, ]

ggplot(top_10_countries, aes(x = reorder(Country, -Number.of.issued), y = Number.of.issued, fill = Country)) +
  geom_bar(stat = "identity") +
  scale_fill_manual(values = rainbow(10)) +
  labs(title = "Top 10 Countries by Number of Visas Issued in 2006",
       x = "Country",
       y = "Number of Visas Issued") +
  theme_minimal() +
  theme(axis.text.x = element_text(angle = 45, hjust = 1)) +
  scale_y_continuous(labels = scales::comma)
```



Example: 2006

```
library(ggplot2)

data_2017 <- subset(data, Year == 2017)

data_2017$Number.of.issued <- as.numeric(data_2017$Number.of.issued)

top_10_countries <- data_2017[order(-data_2017$Number.of.issued), ][1:10, ]
```

```r
ggplot(top_10_countries, aes(x = reorder(Country, -Number.of.issued), y = Number.of.issued, fill = Country)) +
  geom_bar(stat = "identity") +
  scale_fill_manual(values = rainbow(10)) +
  labs(title = "Top 10 Countries by Number of Visas Issued in 2017",
       x = "Country",
       y = "Number of Visas Issued") +
  theme_minimal() +
  theme(axis.text.x = element_text(angle = 45, hjust = 1)) +
  scale_y_continuous(labels = scales::comma)
```



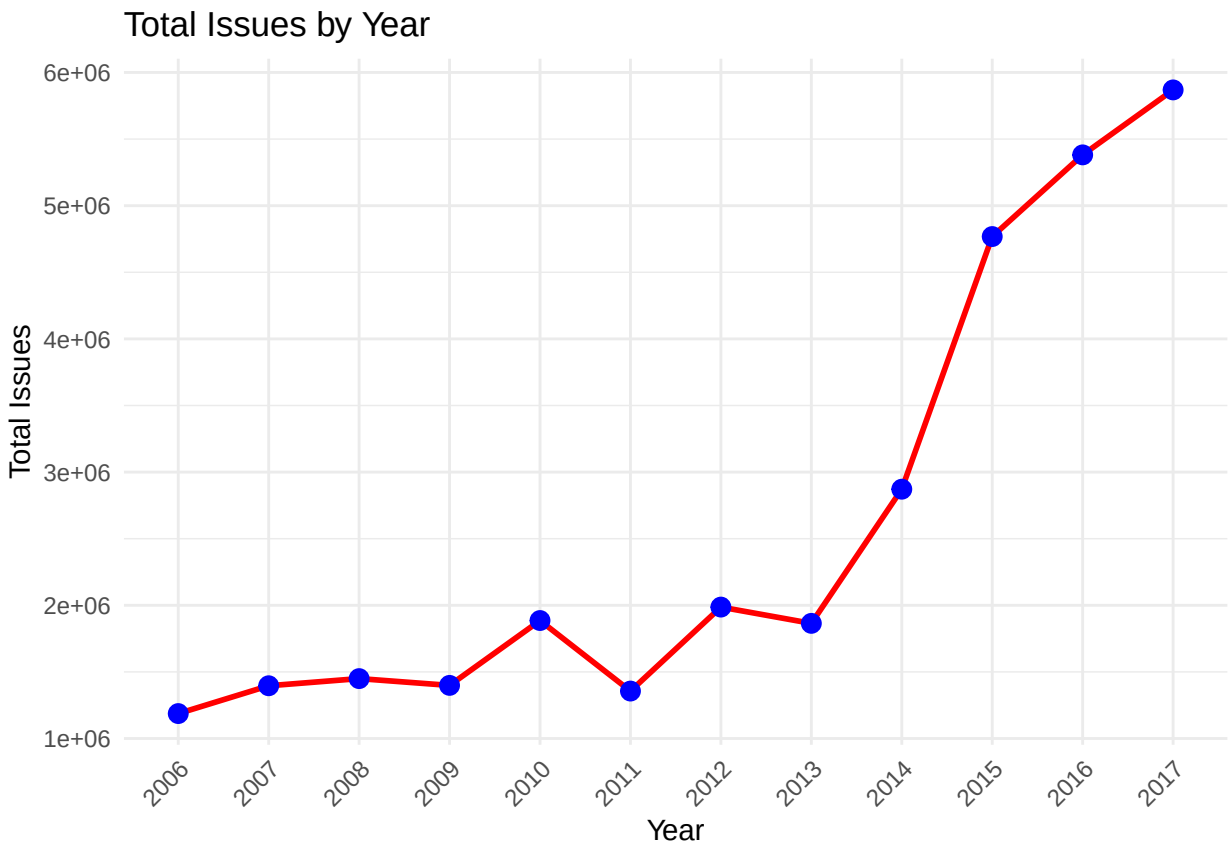
Example: 2017

```
library(dplyr)
```

4.3 The total number of issued for each year through out the dataset

```
##
## Attaching package: 'dplyr'

## The following objects are masked from 'package:stats':
##
##     filter, lag

## The following objects are masked from 'package:base':
##
##     intersect, setdiff, setequal, union
```

```
data_summary <- data %>% group_by(Year) %>% summarise(total_issues = sum(Number.of.issued, na.rm = TRUE))

ggplot(data_summary, aes(x = factor(Year), y = total_issues)) +
  geom_line(group = 1, color = "red", size = 1) +
  geom_point(color = "blue", size = 3) +
  labs(title = "Total Issues by Year", x = "Year", y = "Total Issues") +
  theme_minimal() +
  theme(axis.text.x = element_text(angle = 45, hjust = 1))
```



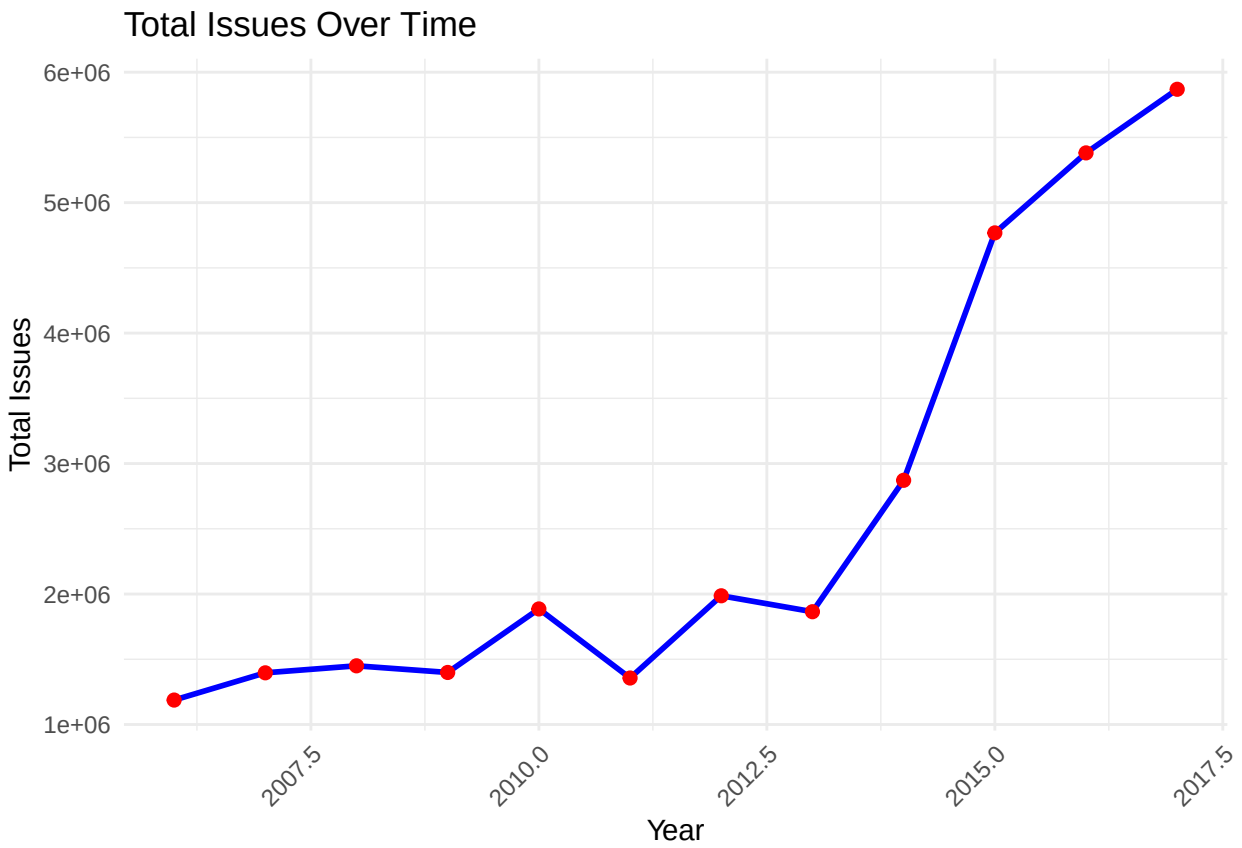
5. Time Series Analysis

```
library(dplyr)

data_summary <- data %>% group_by(Year) %>% summarise(total_issues = sum(Number.of.issued, na.rm = TRUE))

print(data_summary)
```

5.1 Preprocessing

```
## # A tibble: 12 x 2
##     Year total_issues
##    <int>        <int>
##  1  2006      1187218
##  2  2007      1396008
##  3  2008      1449671
##  4  2009      1398756
##  5  2010      1885584
##  6  2011      1356246
##  7  2012      1986539
##  8  2013      1864425
##  9  2014      2871639
## 10  2015      4768286
## 11  2016      5381433
## 12  2017      5869012
```

```
library(ggplot2)

ggplot(data_summary, aes(x = Year, y = total_issues)) +
  geom_line(color = "blue", size = 1) +
  geom_point(color = "red", size = 2) +
  labs(title = "Total Issues Over Time", x = "Year", y = "Total Issues") +
  theme_minimal() +
  theme(axis.text.x = element_text(angle = 45, hjust = 1))
```



5.2 Time Series Plot

```
total_issues_ts <- ts(data_summary$total_issues, start = min(data_summary$Year), frequency = 1)

print(total_issues_ts)
```

5.3 Time Series Decomposition

```
## Time Series:
## Start = 2006
```

```
## End = 2017
## Frequency = 1
##  [1] 1187218 1396008 1449671 1398756 1885584 1356246 1986539 1864425 2871639
## [10] 4768286 5381433 5869012
```

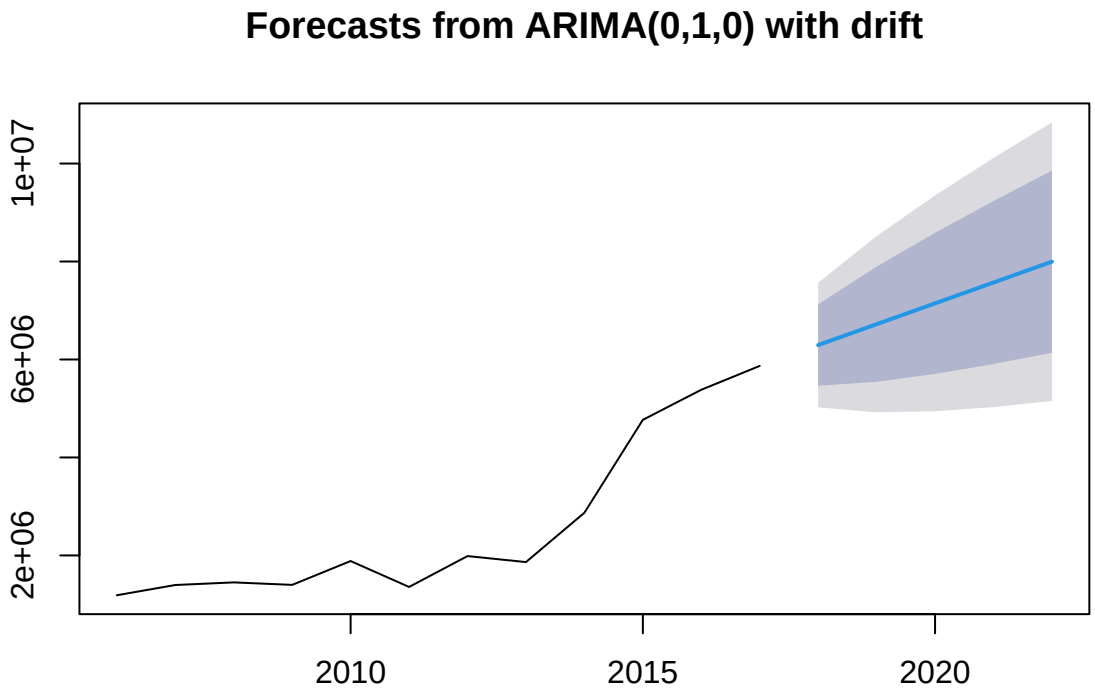
```
library(forecast)
```

5.3 Modeling with ARIMA

```
## Warning: package 'forecast' was built under R version 4.4.2
```

```
## Registered S3 method overwritten by 'quantmod':
##   method            from
##   as.zoo.data.frame zoo
```

```
arima_model <- auto.arima(total_issues_ts)

forecasted_values <- forecast(arima_model, h = 5)

plot(forecasted_values)
```


Forecasts from ARIMA(0,1,0) with drift
1e+07
6e+06
2e+06
2010
2015
2020