

Avances Proyecto 3: Descripcion Variables

Andre Jo 22199 Jose Prince 22087

2025-03-24

```
datos <- read.csv("merged_output.csv")

head(datos)
```

##	DEPREG	MUPREG	MESREG	AÑOREG	DIAOCU	MESOCU	AÑOOCU	DEPOCU	MUPOCU	EDADHOM
## 1	1	101	8	2015	17	7	2015	13	1301	999
## 2	1	101	6	2016	29	7	2015	1	101	999
## 3	19	1901	2	2016	8	7	2015	19	1901	46
## 4	1	101	2	2016	22	12	2015	1	115	999
## 5	22	2204	2	2015	3	2	2015	22	2206	999
## 6	1	101	6	2016	14	10	2015	22	2201	999
##	EDADMUJ	PUEHOM	PUEMUJ	NACHOM	NACMUJ	ESCHOM	ESCMUJ	CIUOHOM	CIUOMUJ	
## 1	999	9	9	320	320	1	9	99.0	23.0	
## 2	999	9	9	840	320	1	9	99.0	41.0	
## 3	40	4	4	320	320	1	9	99.0	98.0	
## 4	999	9	9	320	320	1	9	99.0	98.0	
## 5	999	4	4	320	320	1	9	99.0	99.0	
## 6	999	9	9	320	320	1	9	99.0	99.0	

```
colnames(datos)
```

```
## [1] "DEPREG" "MUPREG" "MESREG" "AÑOREG" "DIAOCU" "MESOCU" "AÑOOCU"
## [8] "DEPOCU" "MUPOCU" "EDADHOM" "EDADMUJ" "PUEHOM" "PUEMUJ" "NACHOM"
## [15] "NACMUJ" "ESCHOM" "ESCMUJ" "CIUOHOM" "CIUOMUJ"
```

Descripción de Variables

Variable Categórica DEPREG - Departamento de registro

Representa el departamento en el cual se registró el evento (1-22)

Variable Categórica MUPREG - Municipio de registro Representa el municipio en el cual se registró el evento (0101-2217)

Variable Numérica MESREG - Mes de registro Representa el mes en el cual se registró el evento (1-12)

Variable Numérica AÑOREG - Año de registro es representado desde (2012-2024)

Variable Numérica DIAOCU - Dia de ocurrencia (1-30)

Variable Numérica MESOCU - Mes de ocurrencia (1-12)

Variable Numérica AÑOOCU - Año ocurrencia (2012-2024)

Variable Categórica DEPOCU - Departamento de Ocurrencia (1-22)

Variable Categórica MUPOCU - Municipio de ocurrencia (0101-2217) Lugar que ocurrió el divorcio

Variable Numérica EDADHOM - Edad del hombre rango de 1-100 año y 999 como ignorado

Variable Numérica EDADMUJ - Edad de la Mujer rango de 1-100 año 999 como ignorado

Variable Categórica PUEHOM - Pueblo de pertenencia del hombre (1-5), 9 is ignorado Es la etnia que pertenece si es Maya, Garifuna o Xinca entre otros.

Variable Categórica PUEMUJ - Pueblo de pertenencia de la mujer (1-5), 9 is ignorado Es la etnia que pertenece si es Maya, Garifuna o Xinca entre otros.

Variable Categórica NACHOM - Nacionalidad del hombre (argentina, peru, Guatemala, Venezuela etc..) Denotados por números aunque se observa que 9999 representa valor ignorado

Variable Categórica NACMUJ - Nacionalidad de la mujer (argentina, peru, Guatemala, Venezuela etc..) Denotados por números aunque se observa que 9999 representa valor ignorado

Variable Categórica ESCHOM - Escolaridad del hombre rango de 1 a 6 siendo 1 ninguno y 6 postgrado 9 es ignorado

Variable Categórica ESCMUJ - Escolaridad de la mujer rango de 1 a 6 siendo 1 ninguno y 6 postgrado 9 es ignorado

Variable Categórica CIUOHOM - Ocupación (Subgrupos, CIUO-08) del hombre Ocupación (trabajo) denotado por 01-96 del otro lado se pone IGNORADO y en caso no estar especificado se pone NEOG

Variable Categórica CIUOMUJ - Ocupación (Subgrupos, CIUO-08) de la mujer Ocupación (trabajo) denotado por 01-96 del otro lado se pone IGNORADO y en caso no estar especificado se pone NEOG

Tomar nota que hay unas discrepancias con las ocupaciones en la lista de variables ya que hay algunos archivos en EXCEL que utilizaron código de 4 dígitos y otros de 2 dígitos. Por otro lado si no se encuentra de la lista se tomara en cuenta como otro etiqueta 97 No especificado en otro grupo Por otro lado si hay 9999 se tomara como IGNORADO

Se observa también que las variables están denotadas por código lo cual se tiene que utilizar excel de definición de variables para saber la etiqueta.

```
length(colnames(datos))
```

```
## [1] 19
```

```
nrow(datos)
```

```
## [1] 81826
```

Se observa que hay un total de 19 variables a medir. De estas existen 7 variables que son numéricas y 12 variables que son categóricas. Además de eso contamos con más de 81826 filas de registros dentro del rango de años 2012 a 2024.

```
summary(datos)
```

```

##          DEPREC          MUPREG          MESREG          AÑOREG
## Min.    : 1.000    Min.    : 101.0    Min.    : 1.000    Min.    :2012
## 1st Qu.: 1.000    1st Qu.: 101.0    1st Qu.: 4.000    1st Qu.:2015
## Median : 6.000    Median : 601.0    Median : 7.000    Median :2019
## Mean    : 7.815    Mean    : 786.5    Mean    : 6.546    Mean    :2018
## 3rd Qu.:13.000    3rd Qu.:1304.0    3rd Qu.:10.000    3rd Qu.:2022
## Max.    :22.000    Max.    :2217.0    Max.    :12.000    Max.    :2024
##
##          DIAOCU          MESOCU          AÑOOCU          DEPOCU
## Min.    : 1.00    Min.    : 1.000    Min.    :2015    Min.    : 1.000
## 1st Qu.: 8.00    1st Qu.: 3.000    1st Qu.:2017    1st Qu.: 1.000
## Median :16.00    Median : 6.000    Median :2020    Median : 6.000
## Mean    :15.78    Mean    : 6.305    Mean    :2020    Mean    : 8.025
## 3rd Qu.:23.00    3rd Qu.: 9.000    3rd Qu.:2022    3rd Qu.:13.000
## Max.    :31.00    Max.    :12.000    Max.    :2023    Max.    :22.000
##
##                                     NA's    :16274
##          MUPOCU          EDADHOM          EDADMUJ          PUEHOM          PUEMUJ
## Min.    : 101    Min.    : 15.0    Min.    : 15.0    Min.    :1.000    Min.    :1.000
## 1st Qu.: 102    1st Qu.: 33.0    1st Qu.: 31.0    1st Qu.:4.000    1st Qu.:4.000
## Median : 613    Median : 54.0    Median : 49.0    Median :9.000    Median :9.000
## Mean    : 808    Mean    :484.1    Mean    :481.5    Mean    :6.279    Mean    :6.303
## 3rd Qu.:1317    3rd Qu.:999.0    3rd Qu.:999.0    3rd Qu.:9.000    3rd Qu.:9.000
## Max.    :2217    Max.    :999.0    Max.    :999.0    Max.    :9.000    Max.    :9.000
##
##          NACHOM          NACMUJ          ESCHOM          ESCMUJ
## Min.    : 10.0    Min.    : 32.0    Min.    :0.000    Min.    :0.00
## 1st Qu.: 320.0    1st Qu.: 320.0    1st Qu.:3.000    1st Qu.:3.00
## Median : 320.0    Median : 320.0    Median :5.000    Median :4.00
## Mean    : 361.6    Mean    : 362.5    Mean    :5.659    Mean    :5.62
## 3rd Qu.: 320.0    3rd Qu.: 320.0    3rd Qu.:9.000    3rd Qu.:9.00
## Max.    :9999.0    Max.    :9999.0    Max.    :9.000    Max.    :9.00
##
##          CIUOHOM          CIUOMUJ
## Length:81826    Length:81826
## Class :character    Class :character
## Mode :character    Mode :character
##
##
##
##

```

Se observa que hay un total de 19 variables a medir. De estas existen 7 variables que son numéricas y 12 variables que son categoricas.

Exploración variables numéricas

Se tienen las siguientes variables como numéricas:

- **AÑOREG** - Año de registro
- **DIAOCU** - Día de ocurrencia
- **MESOCU** - Mes de ocurrencia

- **AÑOOCU** - Año de ocurrencia
- **EDADHOM** - Edad del hombre
- **EDADMUJ** - Edad de la mujer

```
datos <- read.csv("merged_output.csv")

media_EDADHOM <- mean(datos$EDADHOM[datos$EDADHOM != 999], na.rm = TRUE)
media_EDADMUJ <- mean(datos$EDADMUJ[datos$EDADMUJ != 999], na.rm = TRUE)

datos$EDADHOM[datos$EDADHOM == 999] <- media_EDADHOM
datos$EDADMUJ[datos$EDADMUJ == 999] <- media_EDADMUJ

datos_numericos <- datos[, c("AÑOREG", "DIAOCU", "MESOCU", "AÑOOCU", "EDADHOM", "EDADMUJ")]

calcular_estadisticas <- function(columna) {
  lista <- list(
    Media = mean(columna, na.rm = TRUE),
    Mediana = median(columna, na.rm = TRUE),
    Moda = as.numeric(names(sort(table(columna), decreasing = TRUE)[1])),
    Rango = range(columna, na.rm = TRUE),
    Varianza = var(columna, na.rm = TRUE),
    Desviacion_Estandar = sd(columna, na.rm = TRUE),
    Minimo = min(columna, na.rm = TRUE),
    Q1 = quantile(columna, 0.25, na.rm = TRUE),
    Q2 = quantile(columna, 0.50, na.rm = TRUE),
    Q3 = quantile(columna, 0.75, na.rm = TRUE),
    Maximo = max(columna, na.rm = TRUE)
  )
  return(lista)
}

estadisticas <- lapply(datos_numericos, calcular_estadisticas)

for (variable in names(estadisticas)) {
  cat("\n", variable, "\n")
  for (nombre_estadistica in names(estadisticas[[variable]])) {
    cat(nombre_estadistica, ":", estadisticas[[variable]][[nombre_estadistica]], "\n")
  }
}
```

```
##
##  AÑOREG
## Media : 2018.461
## Mediana : 2019
## Moda : 2023
## Rango : 2012 2024
## Varianza : 12.3628
## Desviacion_Estandar : 3.516078
## Minimo : 2012
## Q1 : 2015
## Q2 : 2019
## Q3 : 2022
## Maximo : 2024
##
##  DIAOCU
## Media : 15.77667
## Mediana : 16
## Moda : 22
## Rango : 1 31
## Varianza : 75.38634
## Desviacion_Estandar : 8.68253
## Minimo : 1
## Q1 : 8
## Q2 : 16
## Q3 : 23
## Maximo : 31
##
##  MESOCU
## Media : 6.304646
## Mediana : 6
## Moda : 3
## Rango : 1 12
## Varianza : 10.89804
## Desviacion_Estandar : 3.301217
## Minimo : 1
## Q1 : 3
## Q2 : 6
## Q3 : 9
## Maximo : 12
##
##  AÑOOCU
## Media : 2019.555
## Mediana : 2020
## Moda : 2023
## Rango : 2015 2023
## Varianza : 6.834192
## Desviacion_Estandar : 2.614229
## Minimo : 2015
## Q1 : 2017
## Q2 : 2020
## Q3 : 2022
## Maximo : 2023
```

```
##  
##  EDADHOM  
## Media : 36.31324  
## Mediana : 36.31324  
## Moda : 36.31324  
## Rango : 15 98  
## Varianza : 53.40144  
## Desviacion_Estandar : 7.307629  
## Minimo : 15  
## Q1 : 33  
## Q2 : 36.31324  
## Q3 : 36.31324  
## Maximo : 98  
##  
##  EDADMUJ  
## Media : 33.14645  
## Mediana : 33.14645  
## Moda : 33.14645  
## Rango : 15 81  
## Varianza : 44.62304  
## Desviacion_Estandar : 6.680048  
## Minimo : 15  
## Q1 : 31  
## Q2 : 33.14645  
## Q3 : 33.14645  
## Maximo : 81
```

El análisis de las variables numéricas revela tendencias significativas en los datos. Para las fechas de registro y ocurrencia (AÑOREG y AÑOOCU), se observa que la media y la mediana están cerca de 2019, lo que indica que la mayoría de los eventos se registraron en años recientes. La moda en ambos casos es 2023, lo que sugiere un aumento en los registros en ese año. La dispersión de los datos es baja, con desviaciones estándar de 3.52 y 2.61 respectivamente, lo que implica que la mayoría de los valores se concentran en torno a la media. EN cuanto a los días y meses de ocurrencia (DIAOCU y MESOCU), la distribución parece uniforme, con valores centrales en torno al día 16 y al mes 6. La desviación estándar en el día de ocurrencia es de 8.68, lo que indica mayor variabilidad en los días en comparación con los meses.

Respecto a las edades (EDADHOM y EDADMUJ), las medias son de aproximadamente 36 y 33 años respectivamente, con una dispersión relativamente baja. Esto sugiere que la mayoría de los registros corresponden a adultos jóvenes (este dato debe ser corroborado debido a la omisión de muchas edades en el dataset). El rango de los hombres es mayor (15 - 98 años) en comparación con el de las mujeres (15 - 81 años), lo que podría reflejar una mayor presencia de hombres de mayor edad en los registros. En general, la distribución de las edades es relativamente homogénea.

Exploración variables categoricas

Se tienen las siguientes variables como categóricas:

- **DEPREG** - Departamento de registro
- **MUPREG** - Municipio de registro
- **MESREG** - Mes de registro
- **DEPOCU** - Departamento de ocurrencia

- **MUPOCU** - Municipio de ocurrencia
- **PUEHOM** - Pueblo de pertenencia del hombre
- **PUEMUJ** - Pueblo de pertenencia de la mujer
- **NACHOM** - Nacionalidad del hombre
- **NACMUJ** - Nacionalidad de la mujer
- **ESCHOM** - Escolaridad del hombre
- **ESCMUJ** - Escolaridad de la mujer
- **CIUOHOM** - Ocupación del hombre (según CIUO-08)
- **CIUOMUJ** - Ocupación de la mujer (según CIUO-08)

```
datos_categoricos <- datos[, c("DEPREG", "MUPREG", "DEPOCU", "MUPOCU", "PUEHOM", "PUEMUJ", "NACHOM", "NACMUJ", "ESCHOM", "ESCMUJ", "CIUOHOM", "CIUOMUJ")]
```

```
generar_tabla_frecuencia <- function(columna) {  
  tabla <- table(columna)  
  return(tabla)  
}
```

```
tablas_frecuencia <- head(lapply(datos_categoricos, generar_tabla_frecuencia),10)
```

```
for (variable in names(tablas_frecuencia)) {  
  cat("\n", variable, "\n")  
  print(tablas_frecuencia[[variable]])  
}
```

```

##
##  DEPREG
##  columna
##      1      2      3      4      5      6      7      8      9     10     11     12     13
## 31204 1485 1760 1801 3598 2257  952 1287 6827 2891 2460 3251 2792
##      14     15     16     17     18     19     20     21     22
## 2239 1418 1731 2356 2578 1916 2012 2029 2982
##
##  MUPREG
##  columna
##     101    102    103    104    105    106    107    108    109    110    111    112    113
## 23848   354   345   142   134   589    91   649   196   181   564    56   197
##     114   115   116   117   201   202   203   204   205   206   207   208   301
## 1024 1973  434  427 1175    53    60    21    32    23   104    17   588
##     302   303   304   305   306   307   308   309   310   311   312   313   314
##     170    49    68    39    52    29   255   197    42     8    99    31    81
##     315   316   401   402   403   404   405   406   407   408   409   410   411
##      21     31   919    22   121    46    12    87    72    18    48    37    17
##     412   413   414   415   416   501   502   503   504   505   506   507   508
##      25     75    69   135    98   968   539    85   278   151   472   135    39
##     509   510   511   512   513   514   601   602   603   604   605   606   607
##     295    90   153   121   257    15  1126   163    64    67    18   110     5
##     608   609   610   611   612   613   614   701   704   705   706   707   708
##     247    82    27    98    35    38   177   616    36    38    23     7     1
##     709   710   712   713   714   716   717   718   719   801   802   803   804
##      6     92     5    20     5     1    14    12    76   623   140   118    84
##     805   806   807   808   901   902   903   904   905   906   907   908   909
##     222    60    15    25  3126   172   107   109    40    37    22    17   185
##     910   911   912   913   914   915   916   917   918   919   920   921   922
##      39    77    51    48   166    41   209   246    11    90  1416   201   284
##     923   924  1001  1002  1003  1004  1005  1006  1007  1008  1009  1010  1011
##     100    33  1184   173   263    60    24    87    50   124   111   148    11
##    1012  1013  1014  1015  1016  1017  1018  1019  1020  1021  1101  1102  1103
##      87    56    50    35    44    77    16    50    92   149  1421   142   121
##    1104  1105  1106  1107  1108  1109  1201  1202  1203  1204  1205  1206  1207
##     140   164   107   118   146   101   385   279    48    18    15    15    60
##    1208  1209  1210  1211  1212  1213  1214  1215  1216  1217  1218  1219  1220
##      5     9    55    75    69   158    67   594   144   293   133   101   178
##    1221  1222  1223  1224  1225  1226  1227  1228  1229  1230  1301  1302  1303
##      66   240    40    11    45    10    16    34    29    59  1384   124    78
##    1304  1305  1306  1307  1308  1309  1310  1311  1312  1313  1314  1315  1316
##      61    48    58    84    95    30     4    69   177    19    19    55    20
##    1317  1318  1319  1320  1321  1322  1323  1324  1325  1326  1327  1328  1329
##      48     3    30    16     7    13    17    49    55   116    39     8    23
##    1330  1331  1332  1333  1401  1402  1403  1404  1405  1406  1407  1408  1409
##      6     27     3     7   806    49    51   117    27   130    20    28    22
##    1410  1411  1412  1413  1414  1415  1416  1417  1418  1419  1420  1421  1501
##      29    21   285   173    22    90    87     7    47    23   145    60   675
##    1502  1503  1504  1505  1506  1507  1508  1601  1602  1603  1604  1605  1606
##      81   239   141    58    52   114    58   920   101    82   123    14    12
##    1607  1608  1609  1610  1611  1612  1613  1614  1615  1616  1617  1701  1702
##      16     9   128   124     6    20    39     2   104    20    11   543    31

```



```

## 1703 1704 1705 1706 1707 1708 1709 1710 1711 1712 1713 1714 1801
## 482 63 256 61 51 75 58 48 96 525 48 19 1636
## 1802 1803 1804 1805 1901 1902 1903 1904 1905 1906 1907 1908 1909
## 197 35 538 172 889 221 166 344 117 34 35 27 37
## 1910 1911 2001 2002 2003 2004 2005 2006 2007 2008 2009 2010 2011
## 22 24 804 66 24 36 46 29 408 101 94 86 318
## 2101 2102 2103 2104 2105 2106 2107 2201 2202 2203 2204 2205 2206
## 1899 13 27 17 6 38 29 952 369 209 108 363 37
## 2207 2208 2209 2210 2211 2212 2213 2214 2215 2216 2217
## 111 43 31 15 94 189 32 141 32 131 125
##
## DEPOCU
## columna
## 1 2 3 4 5 6 7 8 9 10 11 12 13
## 29942 1454 1712 1761 3993 2320 968 1489 6255 3060 2432 3598 2926
## 14 15 16 17 18 19 20 21 22
## 2327 1587 1783 2226 2654 1947 2062 1793 3537
##
## MUPOCU
## columna
## 101 102 103 104 105 106 107 108 109 110 111 112 113
## 20197 468 433 105 191 518 175 2478 165 390 169 38 200
## 114 115 116 117 201 202 203 204 205 206 207 208 301
## 981 2251 679 504 516 109 181 54 93 93 327 81 723
## 302 303 304 305 306 307 308 309 310 311 312 313 314
## 141 55 62 12 66 46 208 104 30 13 118 42 47
## 315 316 401 402 403 404 405 406 407 408 409 410 411
## 33 12 659 29 159 83 19 156 87 15 71 23 49
## 412 413 414 415 416 501 502 503 504 505 506 507 508
## 37 100 48 121 105 1213 722 93 113 125 477 209 44
## 509 510 511 512 513 601 602 603 604 605 606 607 608
## 318 78 212 49 340 474 325 134 109 45 109 15 369
## 609 610 611 612 613 614 701 702 703 704 705 706 707
## 177 70 132 61 97 203 363 4 4 91 77 39 21
## 708 709 710 711 712 713 714 715 717 718 719 801 802
## 5 25 149 9 17 58 6 3 23 32 42 731 146
## 803 804 805 806 807 808 901 902 903 904 905 906 907
## 139 91 267 59 22 34 2397 223 168 174 50 52 20
## 908 909 910 911 912 913 914 915 916 917 918 919 920
## 21 234 56 52 57 68 203 51 73 294 20 95 1452
## 921 922 923 924 1001 1002 1003 1004 1005 1006 1007 1008 1009
## 184 102 153 56 1338 278 163 96 41 111 38 131 103
## 1010 1011 1012 1013 1014 1015 1016 1017 1018 1019 1020 1021 1101
## 189 12 38 83 102 62 11 67 30 45 106 16 1273
## 1102 1103 1104 1105 1106 1107 1108 1109 1201 1202 1203 1204 1205
## 185 96 73 160 151 205 156 133 422 382 67 32 21
## 1206 1207 1208 1209 1210 1211 1212 1213 1214 1215 1216 1217 1218
## 31 81 5 20 104 89 94 152 52 606 159 298 191
## 1219 1220 1221 1222 1223 1224 1225 1226 1227 1228 1229 1230 1301
## 139 127 79 165 52 15 81 13 26 33 41 21 1228
## 1302 1303 1304 1305 1306 1307 1308 1309 1310 1311 1312 1313 1314
## 133 61 105 47 55 90 127 40 19 71 247 22 26

```

```

## 1315 1316 1317 1318 1319 1320 1321 1322 1323 1324 1325 1326 1327
## 62 20 52 5 23 20 7 18 25 60 71 159 78
## 1328 1329 1330 1331 1332 1401 1402 1403 1404 1405 1406 1407 1408
## 9 4 9 27 6 670 55 68 119 34 193 17 37
## 1409 1410 1411 1412 1413 1414 1415 1416 1417 1418 1419 1420 1421
## 35 54 24 318 178 30 87 84 9 63 52 126 74
## 1501 1502 1503 1504 1505 1506 1507 1508 1601 1602 1603 1604 1605
## 630 99 238 197 87 93 176 67 896 59 109 124 22
## 1606 1607 1608 1609 1610 1611 1612 1613 1614 1615 1616 1617 1701
## 22 29 21 204 101 9 16 57 5 83 19 7 315
## 1702 1703 1704 1705 1706 1707 1708 1709 1710 1711 1712 1713 1714
## 32 487 80 257 74 40 141 130 97 130 410 21 12
## 1801 1802 1803 1804 1805 1901 1902 1903 1904 1905 1906 1907 1908
## 1447 167 71 670 299 939 148 196 334 97 48 61 34
## 1909 1910 1911 2001 2002 2003 2004 2005 2006 2007 2008 2009 2010
## 46 35 9 914 58 36 70 44 42 346 135 179 34
## 2011 2101 2102 2103 2104 2105 2106 2107 2201 2202 2203 2204 2205
## 204 1135 96 131 62 23 256 90 1306 293 229 140 395
## 2206 2207 2208 2209 2210 2211 2212 2213 2214 2215 2216 2217
## 54 124 34 28 26 108 220 29 194 51 148 158
##
## PUEHOM
## columna
## 1 2 3 4 5 9
## 6291 1143 12 32369 609 41402
##
## PUEMUJ
## columna
## 1 2 3 4 5 9
## 5820 1171 16 32621 687 41511
##
## NACHOM
## columna
## 10 32 36 40 56 68 76 84 100 124 152 156 170
## 2 22 2 5 1 8 9 9 1 18 15 10 51
## 188 192 208 212 214 218 222 276 320 340 344 352 364
## 37 54 2 1 11 12 176 35 80227 64 1 1 2
## 368 376 380 388 392 400 408 410 422 470 484 504 524
## 1 4 17 1 2 5 1 10 3 1 77 1 1
## 554 566 568 578 586 604 608 620 630 642 643 703 710
## 2 1 12 2 1 22 2 1 7 2 1 1 1
## 718 720 723 724 729 752 756 780 818 840 858 862 1000
## 10 87 32 34 1 3 4 1 2 327 4 7 1
## 1020 1021 1026 1034 1043 9999
## 13 6 6 4 2 327
##
## NACMUJ
## columna
## 32 36 40 56 68 76 84 124 152 156 170 188 192
## 15 1 2 3 1 9 4 18 6 4 52 33 50
## 208 214 218 222 276 316 320 340 344 380 392 400 410
## 1 15 6 277 21 1 80155 114 2 4 4 1 1

```

```
## 440 484 498 568 600 604 616 620 630 642 643 718 720
## 1 55 1 4 1 9 1 1 6 1 7 6 116
## 723 724 725 752 756 780 804 826 840 858 862 1020 1026
## 27 13 2 3 2 1 2 1 395 2 12 11 9
## 1034 9999
## 4 333
##
## ESCHOM
## columna
## 0 1 2 3 4 5 6 9
## 457 9660 7936 6334 15951 4542 157 36789
##
## ESCMUJ
## columna
## 0 1 2 3 4 5 6 9
## 542 10067 7622 6275 16452 4401 109 36358
```

El análisis de las tablas de frecuencia de DEPREG y MUPREG revela una distribución desigual de los datos. En DEPREG, la categoría con mayor frecuencia es la primera, con 31,204 casos, lo que representa una concentración significativa en comparación con las demás. Otras categorías muestran una distribución más dispersa, con valores que oscilan entre 952 y 6,827, lo que sugiere una fuerte asimetría en los datos. Este patrón indica que la mayoría de los registros pertenecen a una sola categoría, mientras que las demás están distribuidas en menor medida.

Por otro lado, MUPREG presenta una mayor dispersión con un número más elevado de categorías, algunas con muy pocos registros. La categoría más frecuente (101) tiene 23,848 casos, superando ampliamente a las demás, mientras que muchas otras tienen frecuencias menores a 1,000. Esto sugiere que ciertos valores dominan la distribución mientras que la mayoría son poco representativos. La alta variabilidad en la distribución de frecuencia indica la necesidad de un análisis más profundo para comprender la relevancia de cada categoría y su impacto en el conjunto de datos.

```
datos <- read.csv("merged_output.csv")
```

Se estarán utilizando las siguientes variables para realizar un análisis categorico: “DEPREG”, “MUPREG”, “DEPOCU”, “MUPOCU”, “PUEHOM”, “PUEMUJ”, “NACHOM”, “NACMUJ”, “ESCHOM”, “ESCMUJ”, “CIUOHOM”, “CIUOMUJ”

```
datos_categoricos <- datos[, c("DEPREG", "MUPREG", "DEPOCU", "MUPOCU",
                              "PUEHOM", "PUEMUJ", "NACHOM", "NACMUJ",
                              "ESCHOM", "ESCMUJ", "CIUOHOM", "CIUOMUJ")]

head(datos_categoricos)
```

```
##  DEPREC MUPREG DEPOCU MUPOCU PUEHOM PUEMUJ NACHOM NACMUJ ESCHOM ESCMUJ CIUOHOM
## 1      1    101    13   1301     9     9    320    320     1     9    99.0
## 2      1    101     1    101     9     9    840    320     1     9    99.0
## 3     19   1901    19   1901     4     4    320    320     1     9    99.0
## 4      1    101     1    115     9     9    320    320     1     9    99.0
## 5     22   2204    22   2206     4     4    320    320     1     9    99.0
## 6      1    101    22   2201     9     9    320    320     1     9    99.0
##  CIUOMUJ
## 1     23.0
## 2     41.0
## 3     98.0
## 4     98.0
## 5     99.0
## 6     99.0
```

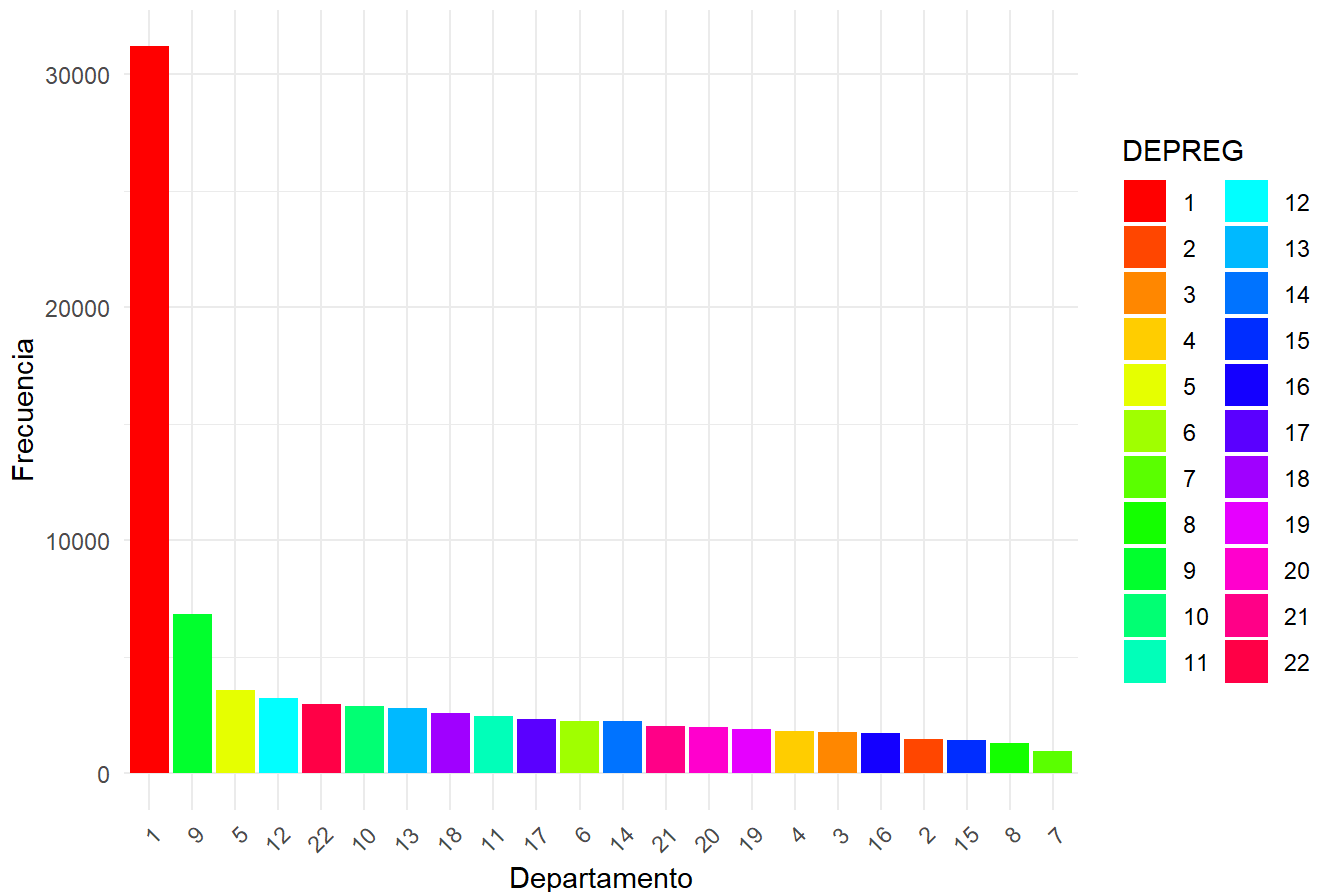
Cuantos registros de divorcios hay en cada departamento?

```
freq_table <- as.data.frame(table(datos$DEPREC))
colnames(freq_table) <- c("DEPREC", "Frecuencia")
print(head(freq_table))
```

```
##  DEPREC Frecuencia
## 1      1    31204
## 2      2     1485
## 3      3     1760
## 4      4     1801
## 5      5     3598
## 6      6     2257
```

```
# Plot with different colors for each category
ggplot(freq_table, aes(x = reorder(DEPREC, -Frecuencia), y = Frecuencia, fill = DEPREC)) +
  geom_bar(stat = "identity") +
  theme_minimal() +
  labs(title = "Frecuencia de Registros por Departamento",
       x = "Departamento",
       y = "Frecuencia") +
  theme(axis.text.x = element_text(angle = 45, hjust = 1)) + # Rotate Labels
  scale_fill_manual(values = rainbow(nrow(freq_table))) # Different colors
```

Frecuencia de Registros por Departamento



1 Guatemala 2 El Progreso 3 Sacatepéquez 4 Chimaltenango 5 Escuintla 6 Santa Rosa 7 Sololá 8 Totonicapán 9 Quetzaltenango 10 Suchitepéquez 11 Retalhuleu 12 San Marcos 13 Huehuetenango 14 Quiché 15 Baja Verapaz 16 Alta Verapaz 17 Petén 18 Izabal 19 Zacapa 20 Chiquimula 21 Jalapa 22 Jutiapa

En base a la gráfica mostrada observamos que el mayor registros de divorcios es en el departamento de Guatemala, En el segundo puesto tenemos a Quetzaltenango con 1485 registros de divorcio, Por último se debe de tomar en cuenta que el departamento de Solola ha marcado el menor registro de divorcios.

Con estos datos posiblemente podemos realizar un analisis ante la relación entre el pueblo o etnia de las personas divorciadas y el departamento donde se registró el divorcio.

```
freq_table <- as.data.frame(table(datos$MUPREG))
colnames(freq_table) <- c("MUPREG", "Frecuencia")
print(head(freq_table))
```

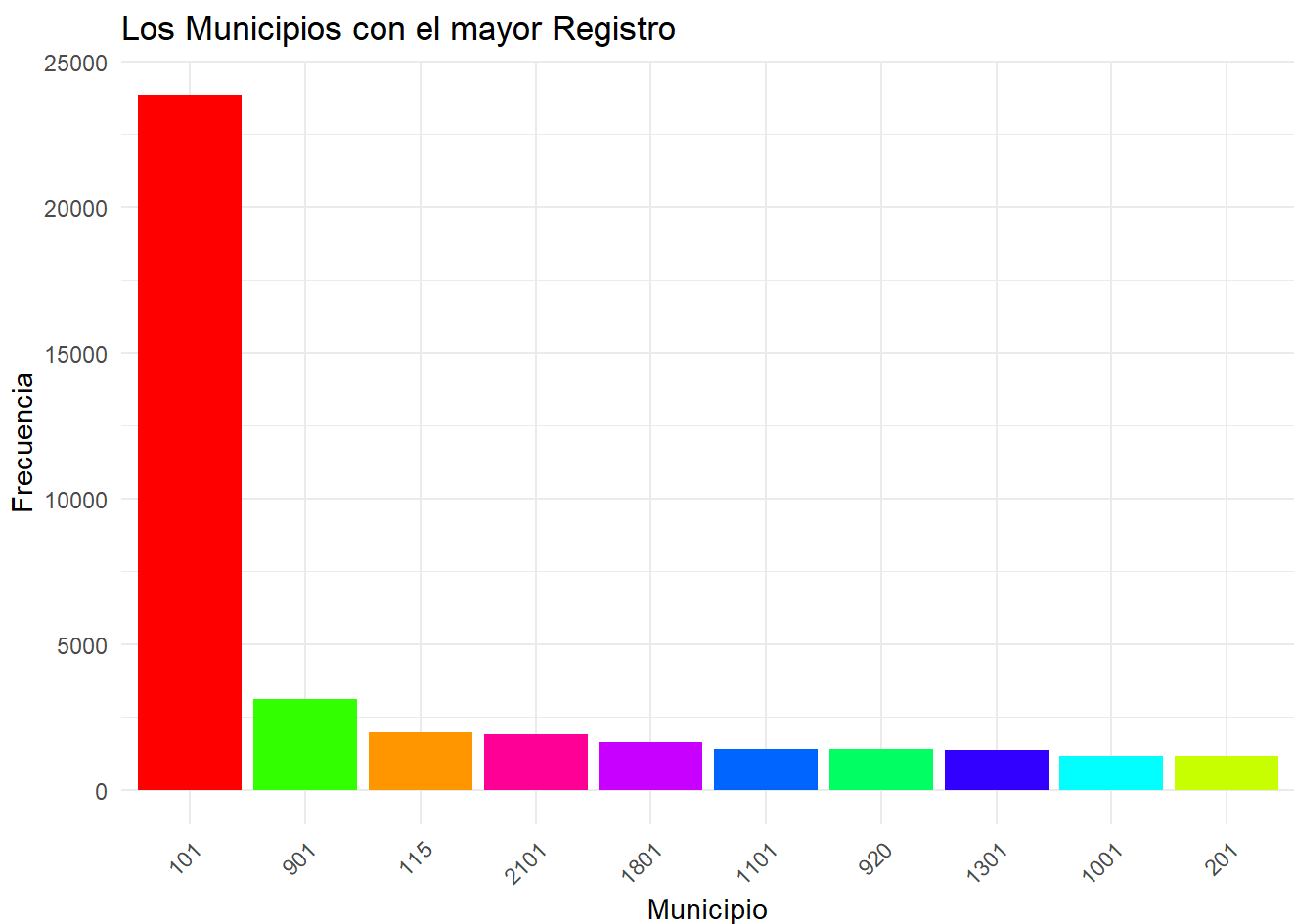
```
##  MUPREG Frecuencia
## 1    101    23848
## 2    102     354
## 3    103     345
## 4    104     142
## 5    105     134
## 6    106     589
```

```

freq_table <- freq_table %>%
  arrange(desc(Frecuencia)) %>%
  head(10)

# Plot with different colors for each category
ggplot(freq_table, aes(x = reorder(MUPREG, -Frecuencia), y = Frecuencia, fill = MUPREG)) +
  geom_bar(stat = "identity") +
  theme_minimal() +
  labs(title = "Los Municipios con el mayor Registro ",
       x = "Municipio",
       y = "Frecuencia") +
  theme(axis.text.x = element_text(angle = 45, hjust = 1), legend.position = "none") + # Rotate
Labels
scale_fill_manual(values = rainbow(nrow(freq_table))) # Different colors

```



Aunque existan una mayor cantidad de municipios se eligieron los 10 municipios con mayor registros. En primer lugar se obtiene el municipio de Guatemala con un monto total de 23848 registros. En segundo lugar tenemos a Santa Catarina Pinula con un total de 354 registros. Se observa que el rango de número es de 101 a 110 lo cual tiene sentido debido a que todos los municipios nombrados forman parte de la capital de Guatemala donde es el departamento con mayor registros según la gráfica anterior.

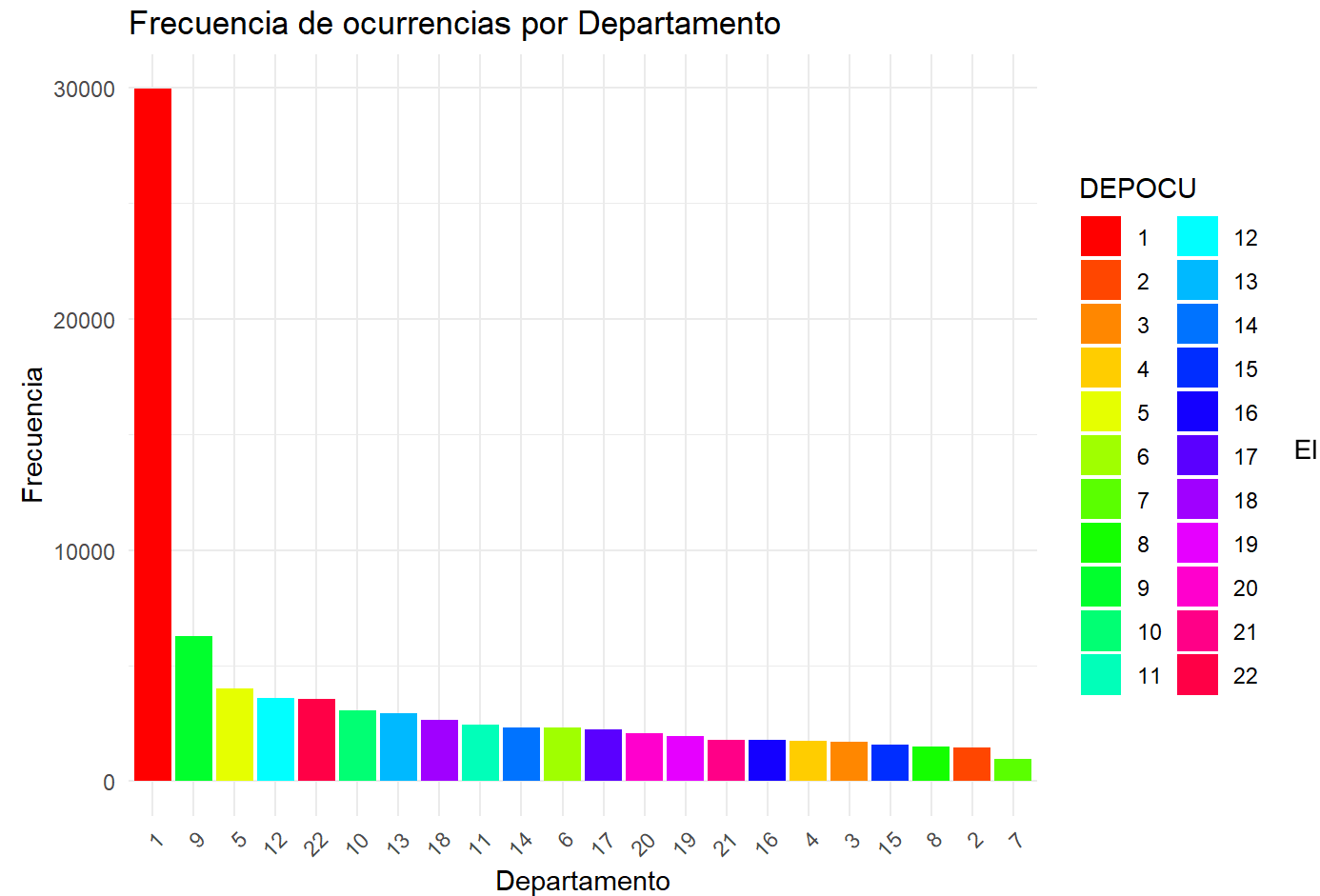
Que departamento es más frecuente en registrar la

ocurrencia del divorcio.

```
freq_table <- as.data.frame(table(datos$DEPOCU))  
colnames(freq_table) <- c("DEPOCU", "Frecuencia")  
print(head(freq_table))
```

```
##   DEPOCU Frecuencia  
## 1      1      29942  
## 2      2       1454  
## 3      3       1712  
## 4      4       1761  
## 5      5       3993  
## 6      6       2320
```

```
# Plot with different colors for each category  
ggplot(freq_table, aes(x = reorder(DEPOCU, -Frecuencia), y = Frecuencia, fill = DEPOCU)) +  
  geom_bar(stat = "identity") +  
  theme_minimal() +  
  labs(title = "Frecuencia de ocurrencias por Departamento",  
        x = "Departamento",  
        y = "Frecuencia") +  
  theme(axis.text.x = element_text(angle = 45, hjust = 1)) + # Rotate Labels  
  scale_fill_manual(values = rainbow(nrow(freq_table))) # Different colors
```



análisis muestra que el departamento de Guatemala tiene la mayor cantidad de registros de divorcios, lo cual es consistente con su alta población y una mayor cantidad de procesos legales en comparación con otros departamentos. Sin embargo, lo interesante es que la diferencia entre el registro y la ocurrencia de los divorcios sugiere que muchas personas pueden haber pensado en divorciarse en el departamento de Guatemala, pero finalmente formalizaron el proceso en otro departamento.

Frecuencia de los municipios con mayor ocurrencia

```
freq_table <- as.data.frame(table(datos$MUPOCU))
colnames(freq_table) <- c("MUPOCU", "Frecuencia")
print(head(freq_table))
```

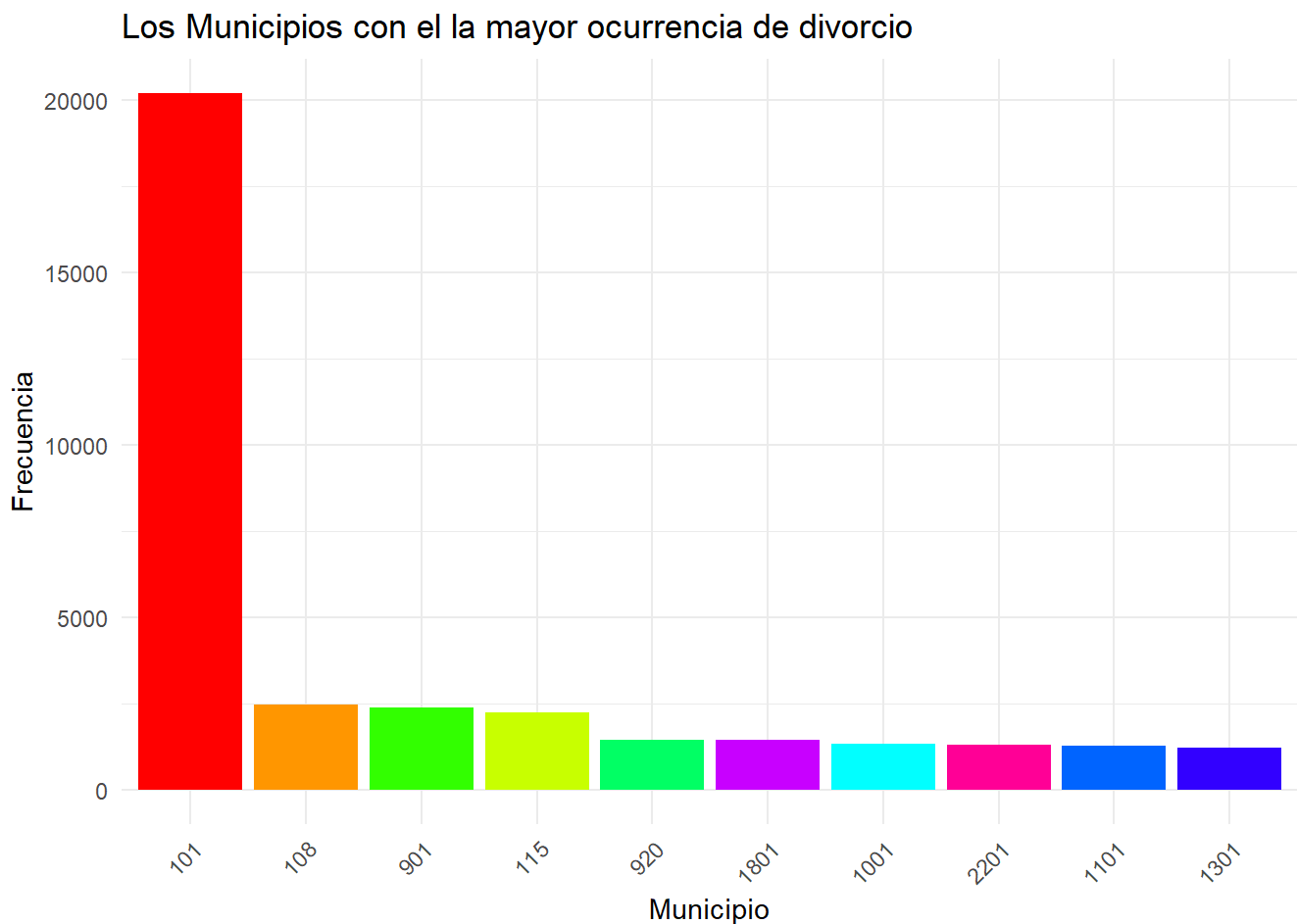
##	MUPOCU	Frecuencia
## 1	101	20197
## 2	102	468
## 3	103	433
## 4	104	105
## 5	105	191
## 6	106	518


```

freq_table <- freq_table %>%
  arrange(desc(Frecuencia)) %>%
  head(10)

# Plot with different colors for each category
ggplot(freq_table, aes(x = reorder(MUPOCU, -Frecuencia), y = Frecuencia, fill = MUPOCU)) +
  geom_bar(stat = "identity") +
  theme_minimal() +
  labs(title = "Los Municipios con el la mayor ocurrencia de divorcio ",
       x = "Municipio",
       y = "Frecuencia") +
  theme(axis.text.x = element_text(angle = 45, hjust = 1), legend.position = "none") + # Rotate
Labels
scale_fill_manual(values = rainbow(nrow(freq_table))) # Different colors

```



Se confirma también cómo, en el departamento, tienen diferentes datos los registros y las ocurrencias, lo cual indica que son variables diferentes. Sin embargo, lo que más se observa es que el municipio de Guatemala tiene la mayor frecuencia en la ocurrencia de divorcio.

Frecuencia de etnias registrados por Hombre

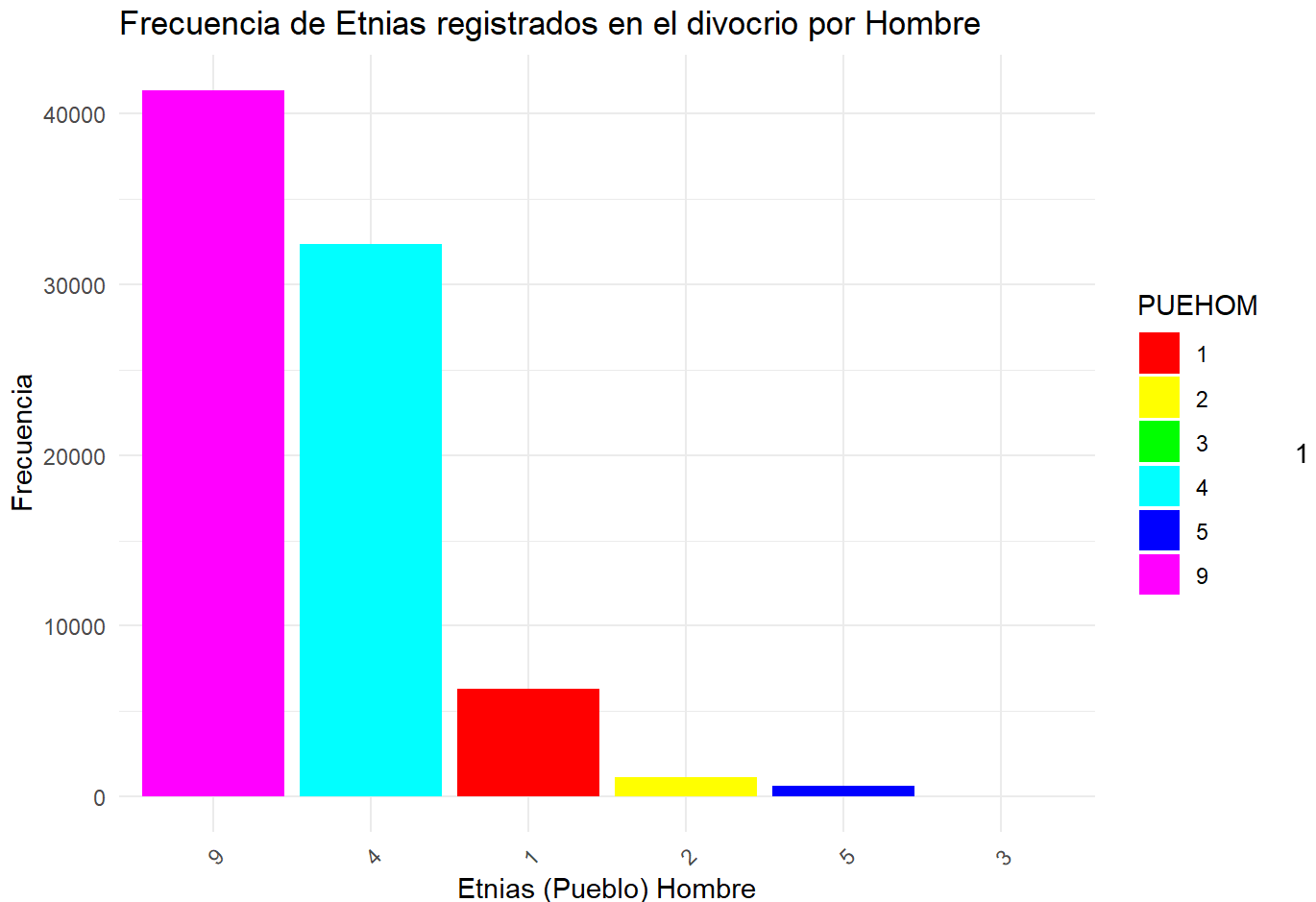
```

freq_table <- as.data.frame(table(datos$PUEHOM))
colnames(freq_table) <- c("PUEHOM", "Frecuencia")
print(freq_table)

```

##	PUEHOM	Frecuencia
## 1	1	6291
## 2	2	1143
## 3	3	12
## 4	4	32369
## 5	5	609
## 6	9	41402

```
# Plot with different colors for each category
ggplot(freq_table, aes(x = reorder(PUEHOM, -Frecuencia), y = Frecuencia, fill = PUEHOM)) +
  geom_bar(stat = "identity") +
  theme_minimal() +
  labs(title = "Frecuencia de Etnias registrados en el divorcio por Hombre",
       x = "Etnias (Pueblo) Hombre",
       y = "Frecuencia") +
  theme(axis.text.x = element_text(angle = 45, hjust = 1)) + # Rotate Labels
  scale_fill_manual(values = rainbow(nrow(freq_table))) # Different colors
```



Maya 2 Garífuna 3 Xinka 4 Mestizo / Ladino 5 Otro 9 Ignorado

Se observa que la mayor frecuencia es IGNORADO aunque observamos en el segundo puesto que los Mestizos/Ladinos con un total de 41402 registros, luego están Maya y la menor frecuencia, se encuentra el pueblo Xinka con un total de 12 registros.

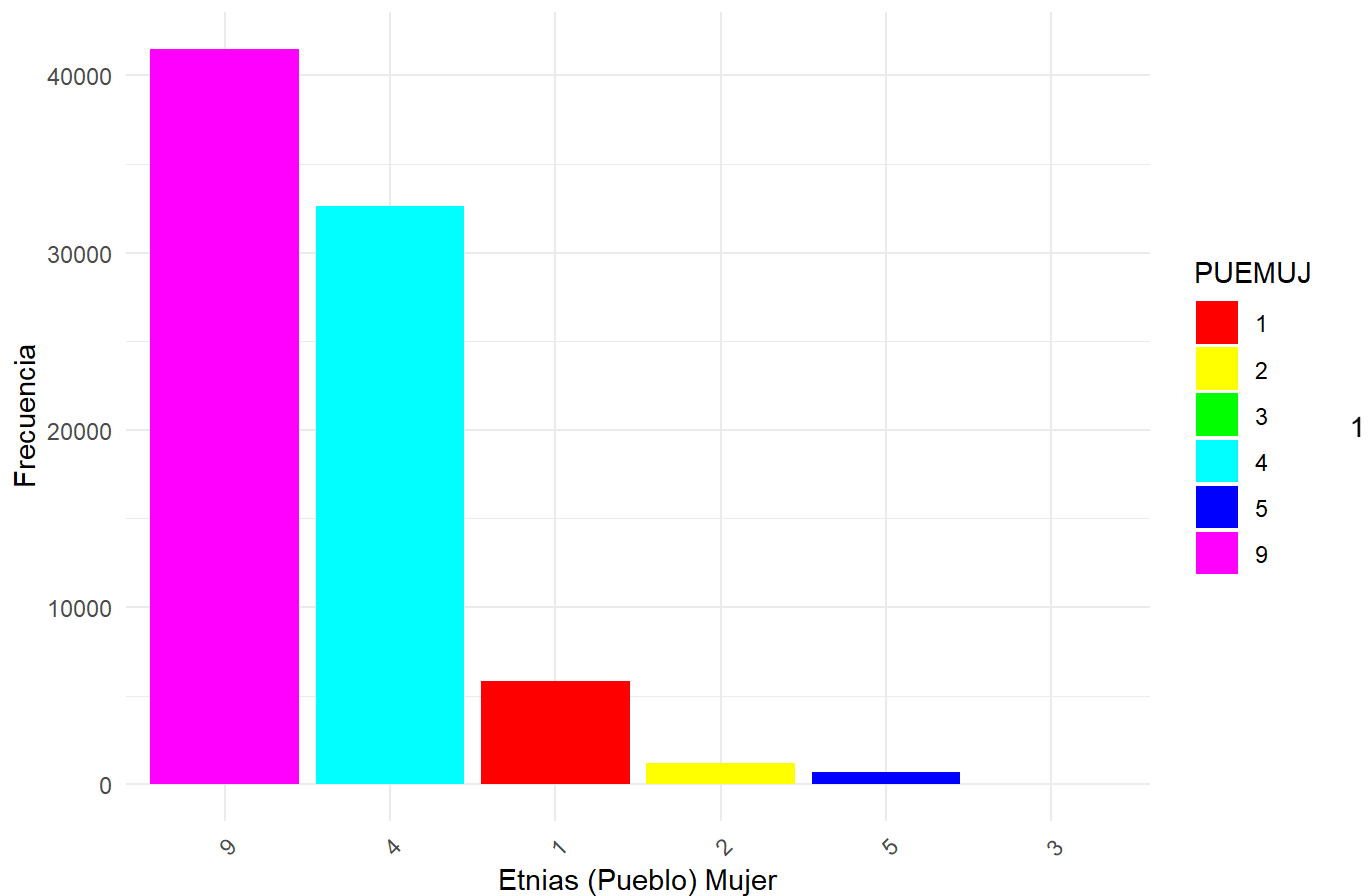
Mujer

```
freq_table <- as.data.frame(table(datos$PUEMUJ))  
colnames(freq_table) <- c("PUEMUJ", "Frecuencia")  
print(freq_table)
```

```
##   PUEMUJ Frecuencia  
## 1      1      5820  
## 2      2      1171  
## 3      3         16  
## 4      4     32621  
## 5      5         687  
## 6      9     41511
```

```
# Plot with different colors for each category  
ggplot(freq_table, aes(x = reorder(PUEMUJ, -Frecuencia), y = Frecuencia, fill = PUEMUJ)) +  
  geom_bar(stat = "identity") +  
  theme_minimal() +  
  labs(title = "Frecuencia de Etnias registrados en el divorcio por Mujer",  
        x = "Etnias (Pueblo) Mujer",  
        y = "Frecuencia") +  
  theme(axis.text.x = element_text(angle = 45, hjust = 1)) + # Rotate Labels  
  scale_fill_manual(values = rainbow(nrow(freq_table))) # Different colors
```

Frecuencia de Etnias registrados en el divorcio por Mujer



Maya 2 Garífuna 3 Xinka 4 Mestizo / Ladino 5 Otro 9 Ignorado

Se observa que la mayor frecuencia es IGNORADO aunque observamos en el segundo puesto que los Mestizos/Ladinos con un total de 32621 registros, luego están Maya y la menor frecuencia, se encuentra el pueblo Xinka con un total de 16 registros.

Nacionalidad de la Hombre

```
freq_table <- as.data.frame(table(datos$NACHOM))
colnames(freq_table) <- c("NACHOM", "Frecuencia")
print(freq_table)
```

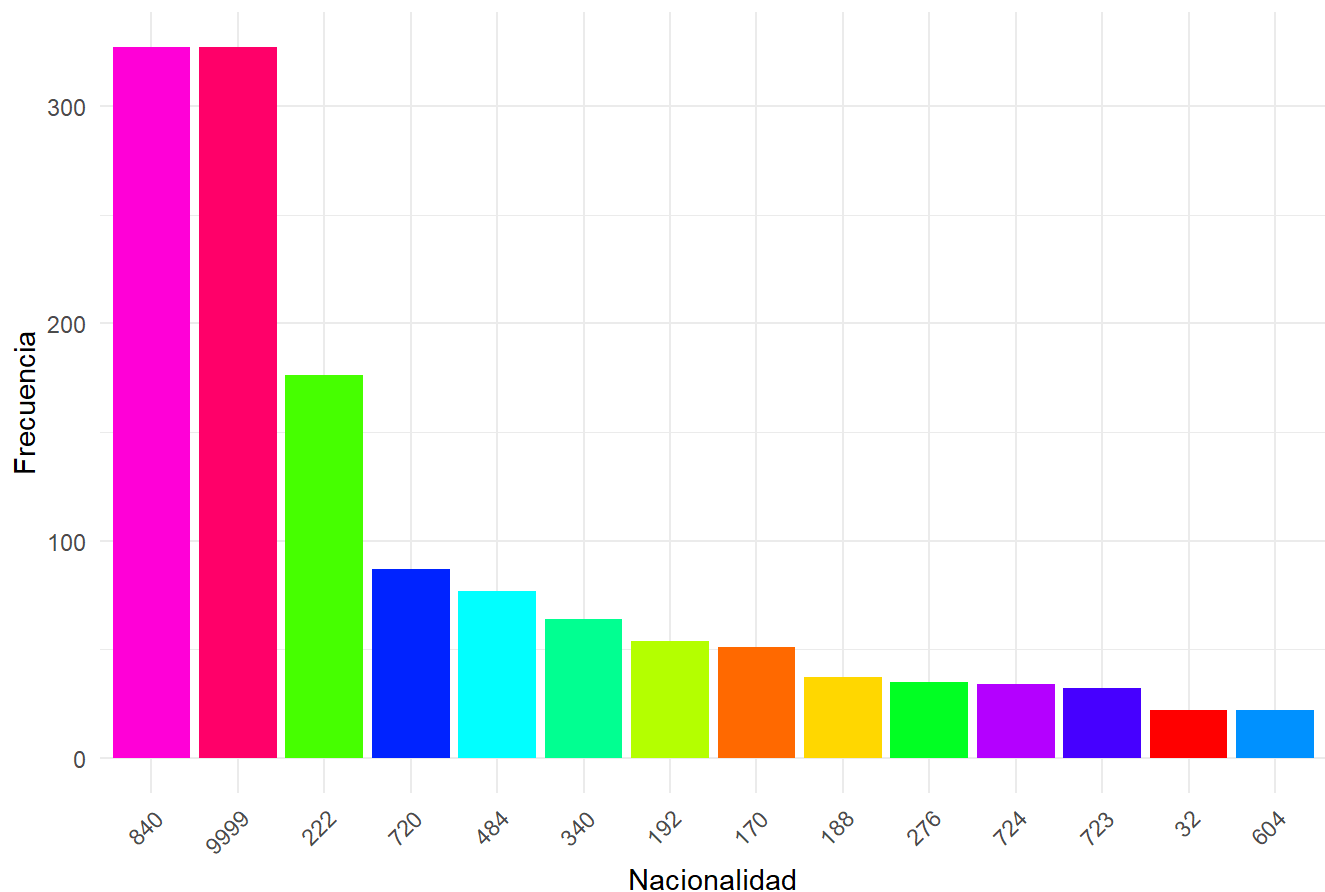
##	NACHOM	Frecuencia
## 1	10	2
## 2	32	22
## 3	36	2
## 4	40	5
## 5	56	1
## 6	68	8
## 7	76	9
## 8	84	9
## 9	100	1
## 10	124	18
## 11	152	15
## 12	156	10
## 13	170	51
## 14	188	37
## 15	192	54
## 16	208	2
## 17	212	1
## 18	214	11
## 19	218	12
## 20	222	176
## 21	276	35
## 22	320	80227
## 23	340	64
## 24	344	1
## 25	352	1
## 26	364	2
## 27	368	1
## 28	376	4
## 29	380	17
## 30	388	1
## 31	392	2
## 32	400	5
## 33	408	1
## 34	410	10
## 35	422	3
## 36	470	1
## 37	484	77
## 38	504	1
## 39	524	1
## 40	554	2
## 41	566	1
## 42	568	12
## 43	578	2
## 44	586	1
## 45	604	22
## 46	608	2
## 47	620	1
## 48	630	7
## 49	642	2
## 50	643	1
## 51	703	1

## 52	710	1
## 53	718	10
## 54	720	87
## 55	723	32
## 56	724	34
## 57	729	1
## 58	752	3
## 59	756	4
## 60	780	1
## 61	818	2
## 62	840	327
## 63	858	4
## 64	862	7
## 65	1000	1
## 66	1020	13
## 67	1021	6
## 68	1026	6
## 69	1034	4
## 70	1043	2
## 71	9999	327

```
freq_table <- freq_table %>%
  arrange(desc(Frecuencia)) %>%
  head(15)
freq_table <- freq_table %>%
  filter(NACHOM != "320") # Removes the row with NACHOM equal to "320"

# Plot with different colors for each category
ggplot(freq_table, aes(x = reorder(NACHOM, -Frecuencia), y = Frecuencia, fill = NACHOM)) +
  geom_bar(stat = "identity") +
  theme_minimal() +
  labs(title = "Frecuencia de nacionalidades por hombre",
       x = "Nacionalidad",
       y = "Frecuencia") +
  theme(axis.text.x = element_text(angle = 45, hjust = 1), legend.position = "none") + # Rotate
Labels
scale_fill_manual(values = rainbow(nrow(freq_table))) # Different colors
```

Frecuencia de nacionalidades por hombre



Tomar nota de que no se agregó la nacionalidad Guatemala debido a que sabemos que tiene un mayor porcentaje. Solo observaremos las de otras nacionalidades para ver cuál es la más frecuente. Se observa que Estados Unidos y la variable IGNORADO son las más frecuentes, ya que tienen 327 datos.

```
freq_table <- as.data.frame(table(datos$NACMUJ))
colnames(freq_table) <- c("NACMUJ", "Frecuencia")
print(head(freq_table))
```

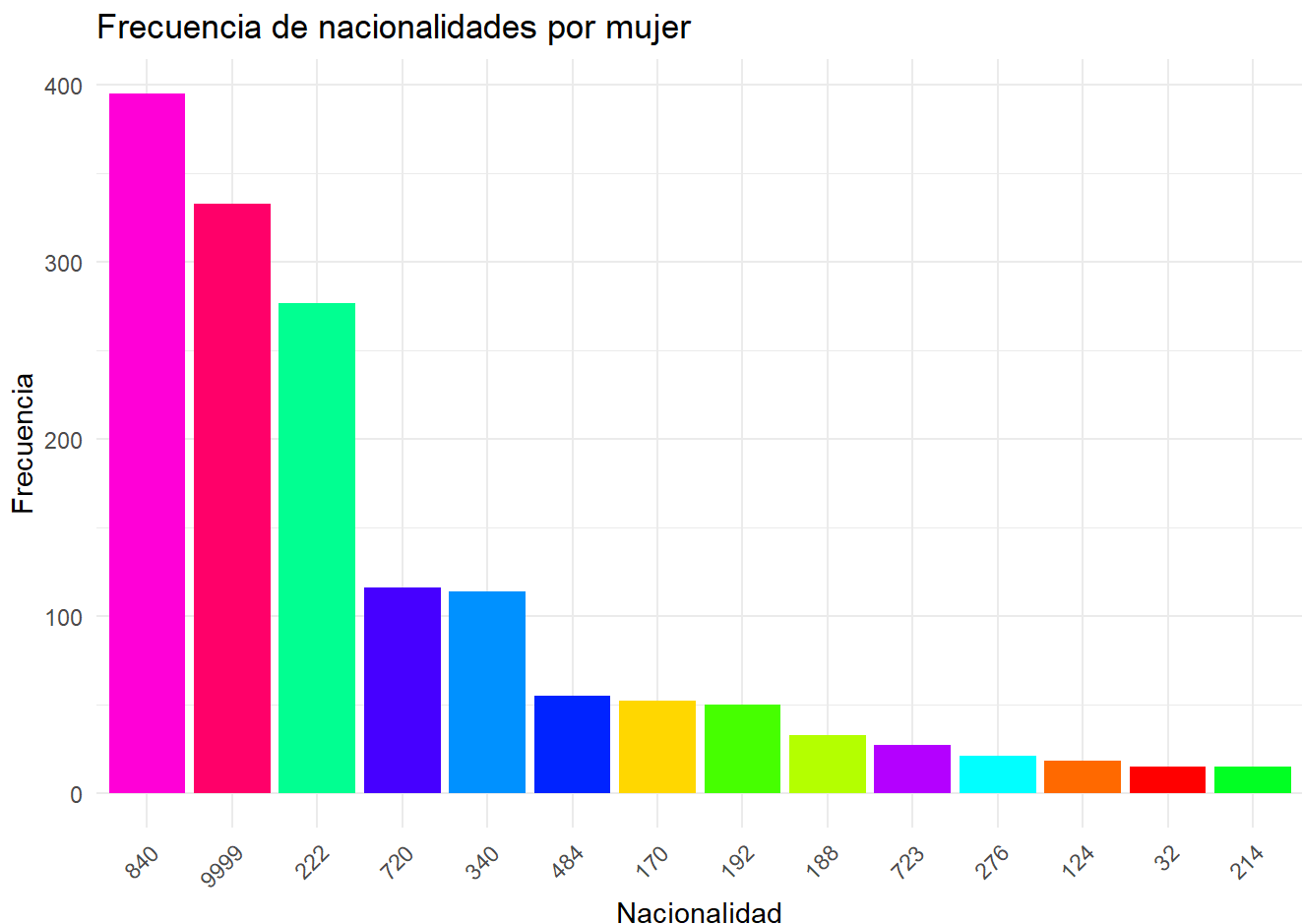
```
##   NACMUJ Frecuencia
## 1     32         15
## 2     36          1
## 3     40          2
## 4     56          3
## 5     68          1
## 6     76          9
```

```

freq_table <- freq_table %>%
  arrange(desc(Frecuencia)) %>%
  head(15)
freq_table <- freq_table %>%
  filter(NACMUJ != "320") # Removes the row with NACHOM equal to "320"

# Plot with different colors for each category
ggplot(freq_table, aes(x = reorder(NACMUJ, -Frecuencia), y = Frecuencia, fill = NACMUJ)) +
  geom_bar(stat = "identity") +
  theme_minimal() +
  labs(title = "Frecuencia de nacionalidades por mujer",
       x = "Nacionalidad",
       y = "Frecuencia") +
  theme(axis.text.x = element_text(angle = 45, hjust = 1), legend.position = "none") + # Rotate
  scale_fill_manual(values = rainbow(nrow(freq_table))) # Different colors
Labels

```



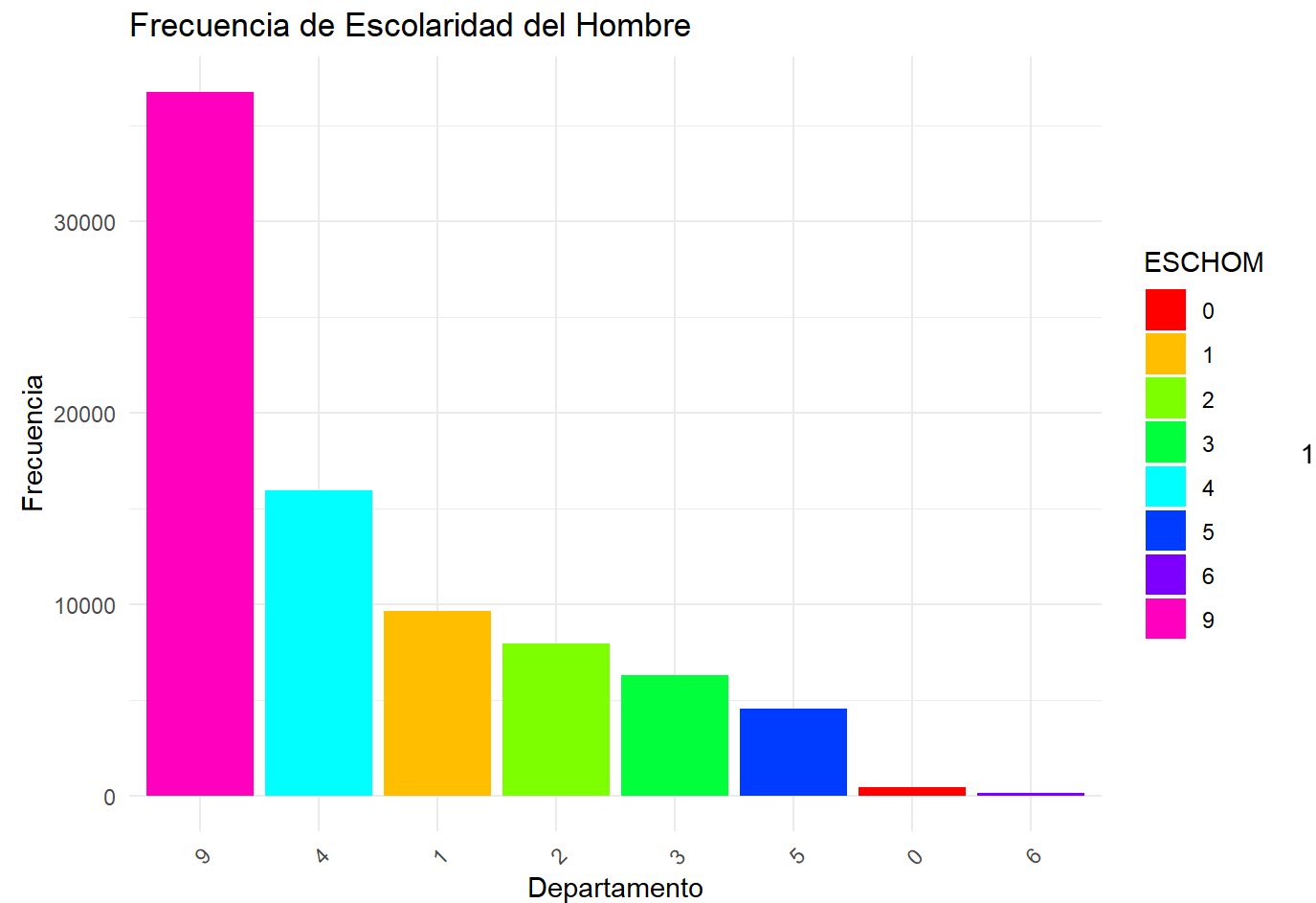
Tomar nota de que no se agregó la nacionalidad Guatemala debido a que sabemos que tiene un mayor porcentaje. Solo observaremos las de otras nacionalidades para ver cuál es la más frecuente. Se observa que Estados Unidos es más frecuente que la variable IGNORADO lo cual hace la mayor frecuente nacionalidad.

Exploración de la variable Escolaridad Hombre

```
freq_table <- as.data.frame(table(datos$ESCHOM))  
colnames(freq_table) <- c("ESCHOM", "Frecuencia")  
print(freq_table)
```

```
##   ESCHOM Frecuencia  
## 1      0         457  
## 2      1        9660  
## 3      2        7936  
## 4      3        6334  
## 5      4       15951  
## 6      5        4542  
## 7      6         157  
## 8      9       36789
```

```
# Plot with different colors for each category  
ggplot(freq_table, aes(x = reorder(ESCHOM, -Frecuencia), y = Frecuencia, fill = ESCHOM)) +  
  geom_bar(stat = "identity") +  
  theme_minimal() +  
  labs(title = "Frecuencia de Escolaridad del Hombre",  
        x = "Departamento",  
        y = "Frecuencia") +  
  theme(axis.text.x = element_text(angle = 45, hjust = 1)) + # Rotate Labels  
  scale_fill_manual(values = rainbow(nrow(freq_table))) # Different colors
```



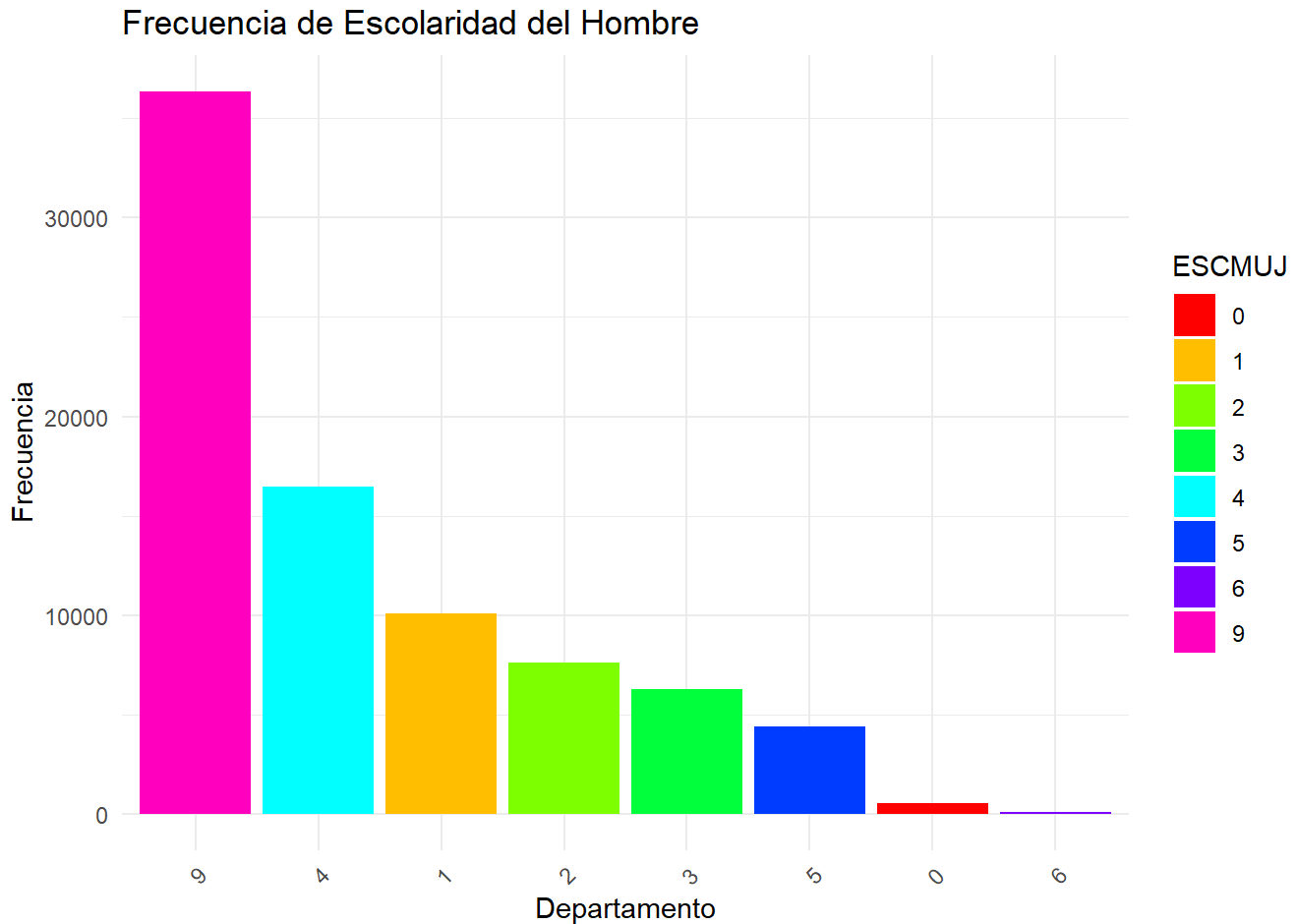
Ninguno 2 Primaria 3 Básica 4 Diversificado 5 Universitario 6 Postgrado 9 Ignorado

Se observa que la mayor frecuencia corresponde a la variable IGNORADO, lo cual no nos indica si la población está educada o no. Por otro lado, observamos que la variable ‘diversificado’ (15,951) es la más frecuente y tiene cierta significancia. Otro dato importante a resaltar es que la variable ‘ninguno’ es la tercera más frecuente, lo cual podríamos relacionar con la causa del divorcio, si esta está vinculada con la educación.

```
freq_table <- as.data.frame(table(datos$ESCMUJ))
colnames(freq_table) <- c("ESCMUJ", "Frecuencia")
print(freq_table)
```

##	ESCMUJ	Frecuencia
## 1	0	542
## 2	1	10067
## 3	2	7622
## 4	3	6275
## 5	4	16452
## 6	5	4401
## 7	6	109
## 8	9	36358

```
# Plot with different colors for each category
ggplot(freq_table, aes(x = reorder(ESCMUJ, -Frecuencia), y = Frecuencia, fill = ESCMUJ)) +
  geom_bar(stat = "identity") +
  theme_minimal() +
  labs(title = "Frecuencia de Escolaridad del Hombre",
       x = "Departamento",
       y = "Frecuencia") +
  theme(axis.text.x = element_text(angle = 45, hjust = 1)) + # Rotate Labels
  scale_fill_manual(values = rainbow(nrow(freq_table))) # Different colors
```



Se observa nuevamente que la mayor frecuencia corresponde a la variable IGNORADO, lo cual no nos indica si la población está educada o no. Por otro lado, observamos que la variable 'diversificado' (15,951) es la más frecuente y tiene cierta significancia. Otro dato importante a resaltar es que la variable 'ninguno' es la tercera más frecuente, lo cual podríamos relacionar con la causa del divorcio, si esta está vinculada con la educación.

Ocupacion Hombres

```
replace_with_median <- function(df, column_name) {  
  # Extract the first two digits from the specified column  
  df <- df %>%  
    mutate(!!column_name := as.integer(substr(as.character(!!sym(column_name)), 1, 2)))  
  
  # Calculate the median of the transformed column  
  median_value <- median(df[[column_name]], na.rm = TRUE)  
  
  # Replace NA values with the median  
  df[[column_name]][is.na(df[[column_name]])] <- median_value  
  
  return(df)  
}
```

```
datos <- replace_with_median(datos, "CIUOHOM")  
datos <- replace_with_median(datos, "CIUOMUJ")
```

```
freq_table <- as.data.frame(table(datos$CIUOHOM))  
colnames(freq_table) <- c("CIUOHOM", "Frecuencia")  
print(freq_table)
```

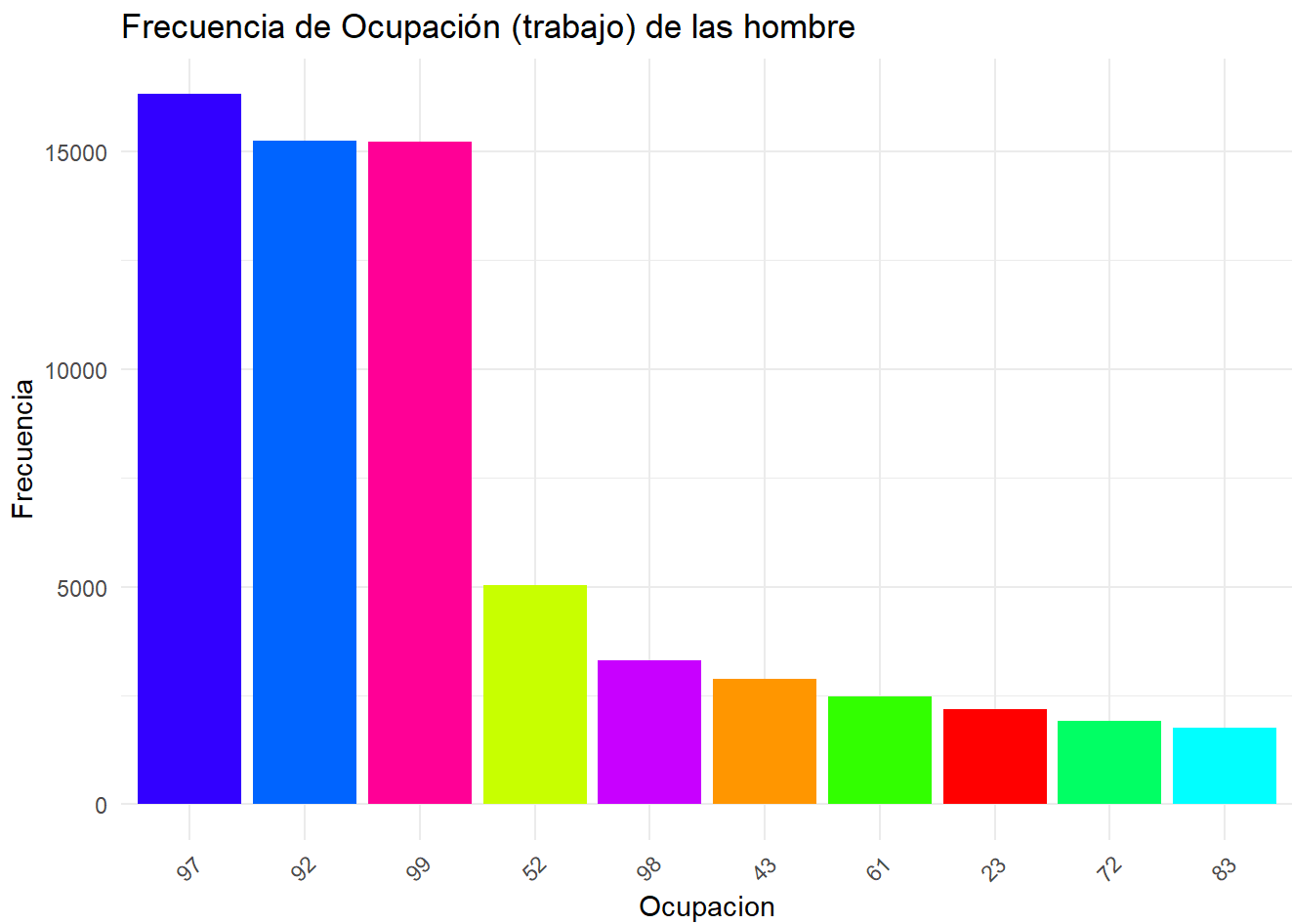
##	CIUOHOM	Frecuencia
## 1	1	88
## 2	3	82
## 3	11	77
## 4	12	306
## 5	13	41
## 6	14	8
## 7	21	1195
## 8	22	602
## 9	23	2192
## 10	24	618
## 11	25	252
## 12	26	855
## 13	31	1581
## 14	32	208
## 15	33	1002
## 16	34	1317
## 17	35	1519
## 18	41	155
## 19	42	226
## 20	43	2881
## 21	44	85
## 22	51	484
## 23	52	5040
## 24	53	8
## 25	54	433
## 26	61	2482
## 27	62	26
## 28	71	1696
## 29	72	1922
## 30	73	244
## 31	74	436
## 32	75	745
## 33	81	626
## 34	82	7
## 35	83	1751
## 36	91	89
## 37	92	15251
## 38	93	279
## 39	94	5
## 40	95	11
## 41	96	143
## 42	97	16316
## 43	98	3317
## 44	99	15225

```

freq_table <- freq_table %>%
  arrange(desc(Frecuencia)) %>%
  head(10)

# Plot with different colors for each category
ggplot(freq_table, aes(x = reorder(CIUOHOM, -Frecuencia), y = Frecuencia, fill = CIUOHOM)) +
  geom_bar(stat = "identity") +
  theme_minimal() +
  labs(title = "Frecuencia de Ocupación (trabajo) de las hombre ",
       x = "Ocupacion",
       y = "Frecuencia") +
  theme(axis.text.x = element_text(angle = 45, hjust = 1), legend.position = "none") + # Rotate
Labels
scale_fill_manual(values = rainbow(nrow(freq_table))) # Different colors

```



```

freq_table <- as.data.frame(table(datos$CIUOMUJ))
colnames(freq_table) <- c("CIUOMUJ", "Frecuencia")
print(freq_table)

```

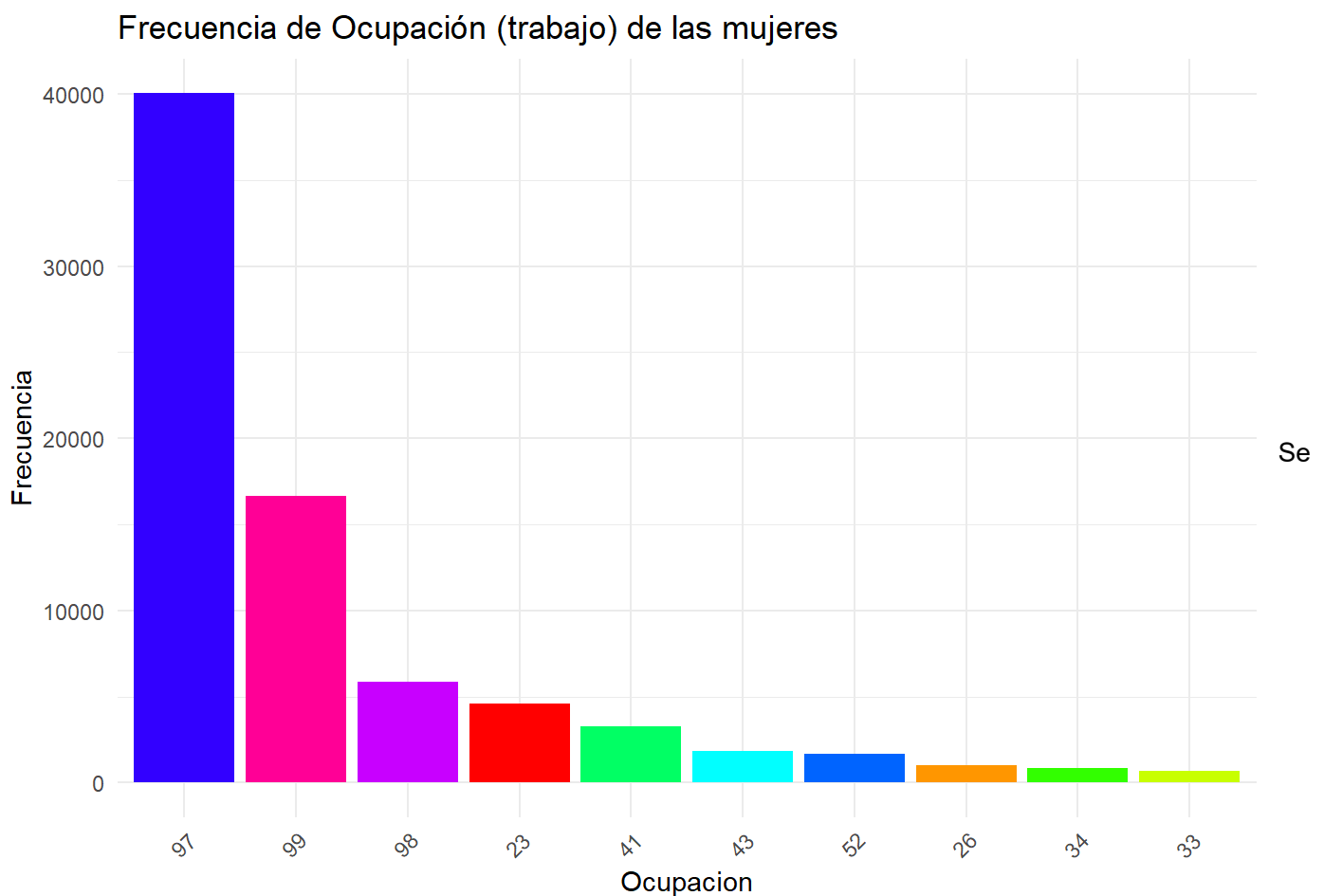
##	CIUOMUJ	Frecuencia
## 1	1	9
## 2	3	5
## 3	11	40
## 4	12	225
## 5	13	5
## 6	14	14
## 7	21	479
## 8	22	613
## 9	23	4591
## 10	24	533
## 11	25	55
## 12	26	1036
## 13	31	145
## 14	32	547
## 15	33	678
## 16	34	857
## 17	35	588
## 18	41	3293
## 19	42	275
## 20	43	1865
## 21	44	134
## 22	51	615
## 23	52	1682
## 24	53	71
## 25	54	31
## 26	61	2
## 27	71	5
## 28	72	6
## 29	73	54
## 30	75	387
## 31	81	254
## 32	82	1
## 33	83	6
## 34	91	80
## 35	92	18
## 36	93	41
## 37	94	9
## 38	95	2
## 39	96	21
## 40	97	40056
## 41	98	5855
## 42	99	16643

```

freq_table <- freq_table %>%
  arrange(desc(Frecuencia)) %>%
  head(10)

# Plot with different colors for each category
ggplot(freq_table, aes(x = reorder(CIUOMUJ, -Frecuencia), y = Frecuencia, fill = CIUOMUJ)) +
  geom_bar(stat = "identity") +
  theme_minimal() +
  labs(title = "Frecuencia de Ocupación (trabajo) de las mujeres ",
       x = "Ocupacion",
       y = "Frecuencia") +
  theme(axis.text.x = element_text(angle = 45, hjust = 1), legend.position = "none") + # Rotate
Labels
scale_fill_manual(values = rainbow(nrow(freq_table))) # Different colors

```



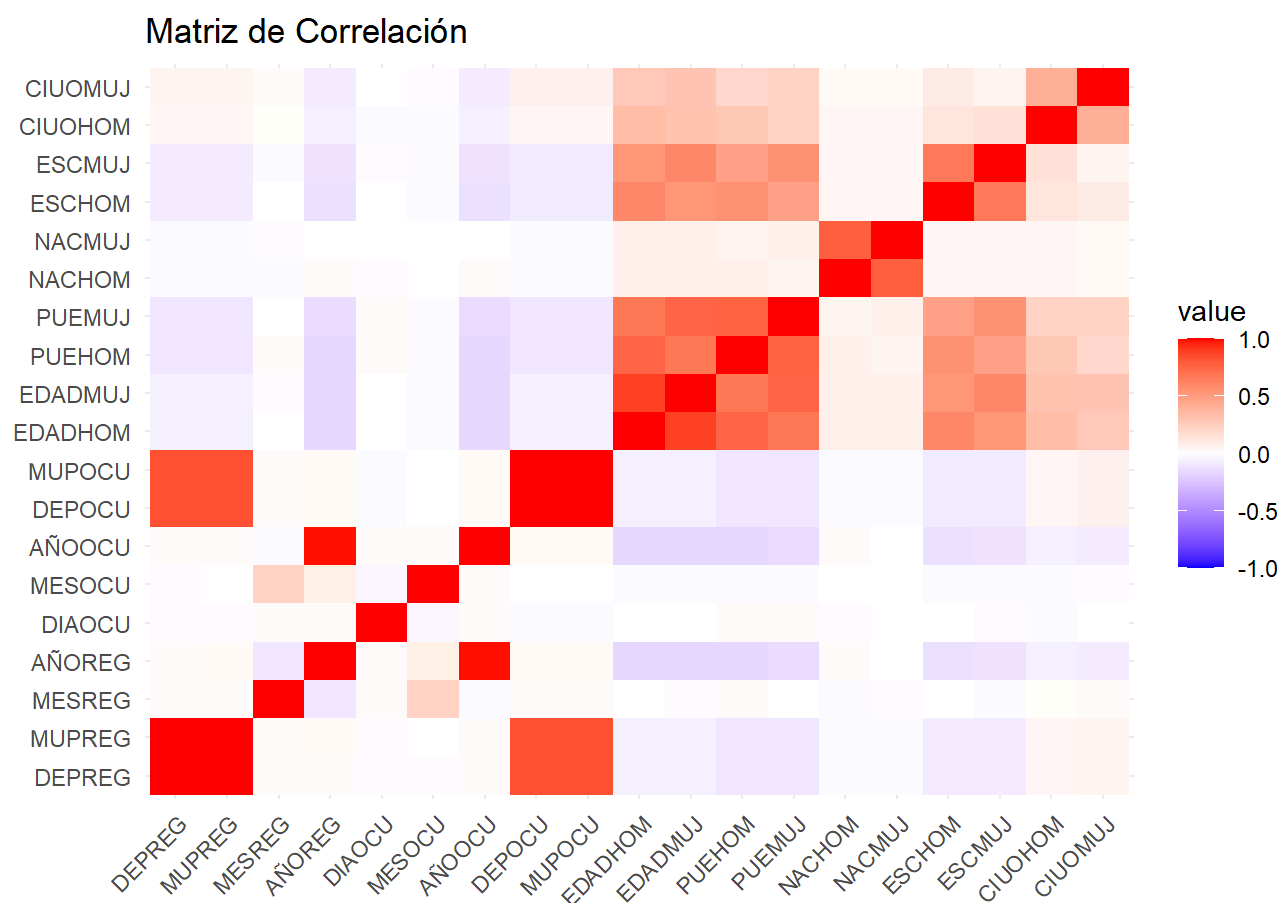
observa que la variable con mayor frecuencia es “No especificado en otro grupo”, lo cual puede representar cualquier ocupación no mencionada. Por otro lado, observamos que 99 y 98 corresponden a otra variable que no tiene gran significancia, ya que se clasifica como “Ignorado”. Además, se puede notar que el dato 23 tiene relevancia, ya que representa ocupaciones de peones agropecuarios, pesqueros y forestales, siendo uno de los datos más frecuentes.

Relaciones entre las variables


```
numerical_data <- datos[sapply(datos, is.numeric)]

corr_matrix <- cor(numerical_data, use = "complete.obs")
corr_melted <- melt(corr_matrix)

ggplot(corr_melted, aes(Var1, Var2, fill = value)) +
  geom_tile() +
  scale_fill_gradient2(low = "blue", high = "red", mid = "white", midpoint = 0, limit = c(-1,
1)) +
  theme_minimal() +
  labs(title = "Matriz de Correlación", x = "", y = "") +
  theme(axis.text.x = element_text(angle = 45, hjust = 1))
```



La matriz de correlación muestra relaciones lineales entre las variables, se revelan fuertes correlaciones positivas entre fechas de registro y ocurrencia, así como entre las edades de hombres y mujeres. Sin embargo, las variables categóricas como pueblos, nacionalidades y escolaridad muestran correlaciones débiles o nulas entre sí y con las variables de fecha y edad, indicando que no hay una relación lineal fuerte.