



Dia 3. Proyectos

Tema 6. Análisis estadístico de datos geográficos en R

Proyecto 1.

PROFACAD: Producto 6 para el portafolio

Producto para el portafolio: Producto 6. Proyecto 1

Fecha de entrega del producto: 26 de julio 2019

Actividades: En equipos de 2 alumnos realizar una síntesis estadística de los datos sobre población por municipio en un estado por elegir, utilizando datos del Censo de Población 2010 y representarlos en forma gráfica y cartográfica (proyecto 1).

Producto para el portafolio: Reporte de análisis de datos sobre población por municipio en el estado elegido en formato del documento R Markdown PDF con los fragmentos de código insertados, gráficas y productos de visualización, que cumpla con las características definidas en los lineamientos del proyecto 1.

PROFACAD: Producto 6 para el portafolio

Producto para el portafolio: Producto 6. Proyecto 1

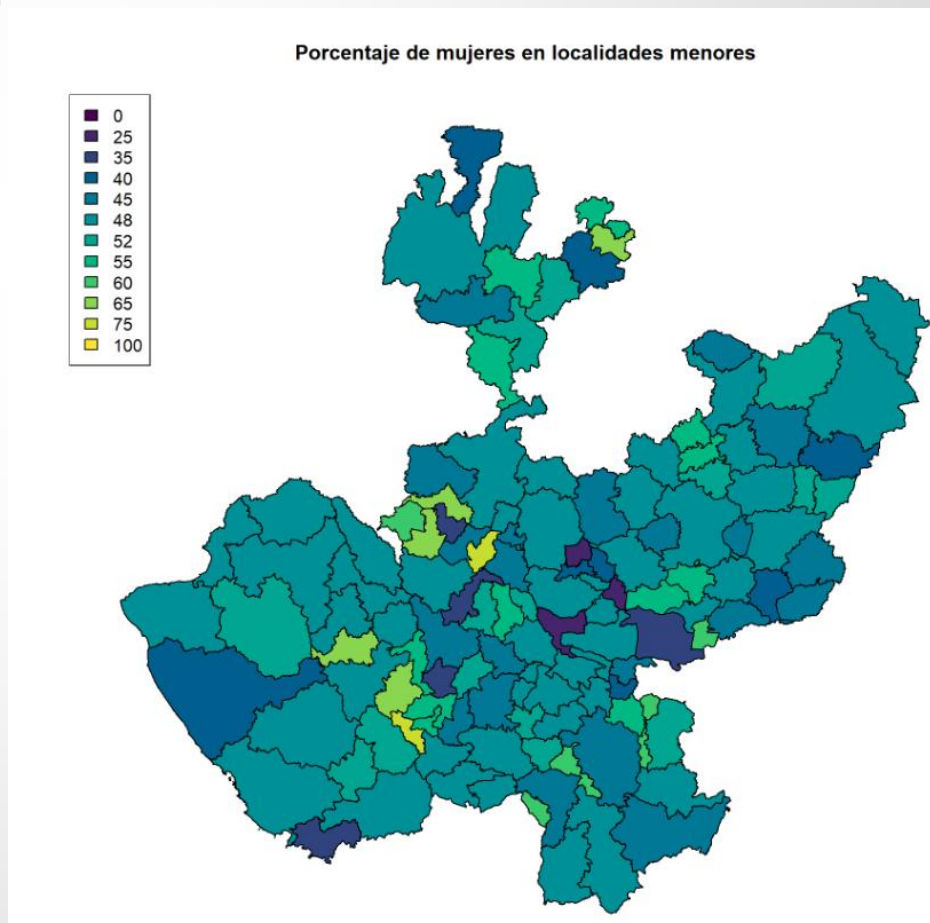
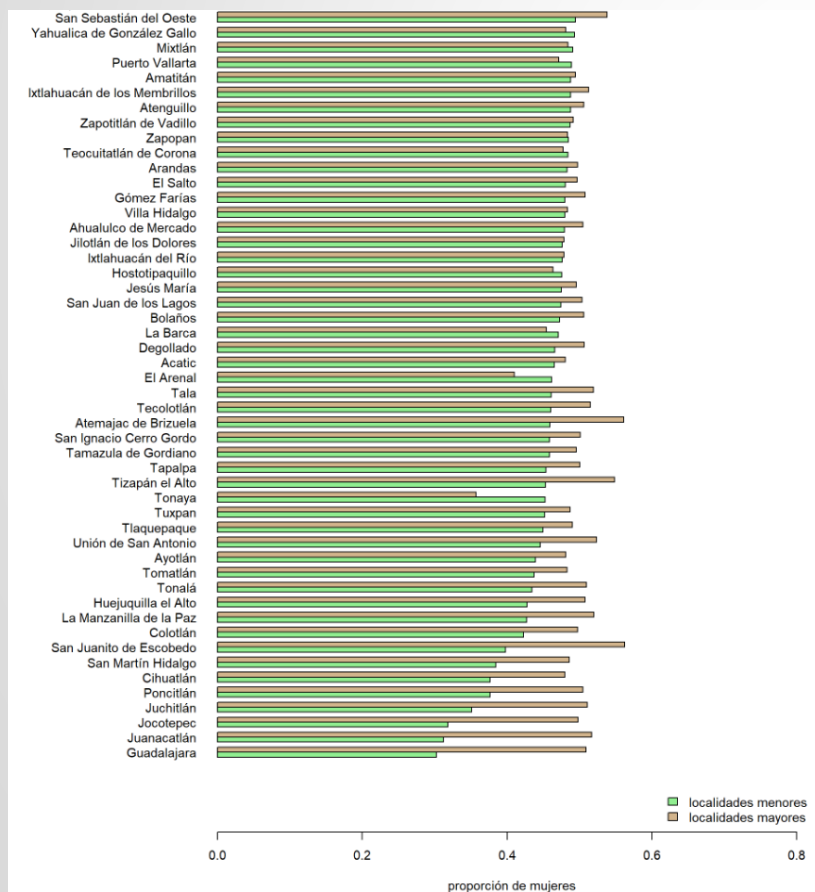
Instrucciones:

- Utilizar datos del Censo de Población 2010 para un estado (a elegir)
- Calcular la proporción de hombres y mujeres por municipio del estado, por separado en el conjunto de localidades menores (con menos que 50 habitantes) y en localidades mayores.
- Visualizar los resultados en forma cartográfica (obligatorio) y gráfica (opcional). Se sugiere generar dos mapas con municipios coloreados de acuerdo con proporción de mujeres en total de población, un mapa considerando localidades menores, y otra localidades mayores. Tipo de grafico recomendado – grafica de barras por municipio con dos variables.

PROFACAD: Producto 6 para el portafolio


Producto para el portafolio: Producto 6. Proyecto 1

Ejemplos de grafica (fragmento) y de mapa



Proyecto 1 – datos fuente

<https://www.inegi.org.mx/programas/ccpv/2010/default.html>

**INSTITUTO NACIONAL
DE ESTADÍSTICA Y GEOGRAFÍA**

Censo de Población y Vivienda 2010

Seleccionar una entidad federativa

☐ Distrito Federal

☐ Durango

☐ Guanajuato

☐ Guerrero

☐ Hidalgo

☒ Jalisco

DBF

XLS

TXT

Censo de Población y Vivienda 2010

Cambiar evento

Descripción de datos

Manual de usuario

Nueva consulta

11 314 Localidades encontradas

227 páginas (50 registros por página)

ANAL Y GEOGRAFÍA

enda 2010

federativa

DBF XLS TXT

Municipio o delegación		Localidad	
Clave	Nombre	Clave	Nombre
001	Acatic	9998	Localidades de una vivienda
001	Acatic	9999	Localidades de dos viviendas
001	Acatic	0001	Acatic
001	Acatic	0002	Agua Prieta (Tepetates)
001	Acatic	0003	Alacates
001	Acatic	0006	El Bajío de los Herrera
001	Acatic	0009	Rancho Nuevo Calderón
001	Acatic	0010	El Capadero (San Miguel de la Loza)
001	Acatic	0011	Carricillo
001	Acatic	0012	Ceja de Ramblas
001	Acatic	0013	Cofradía Grande
001	Acatic	0014	El Colomo
001	Acatic	0016	Las Coronillas
001	Acatic	0018	La Chorrera

Notas: * Datos reservados por confidencialidad N/D No disponible N/A No aplica

Formato

Primero « 1 2 3 4 5 » Último

Regresar

Proyecto 1 (revisar datos)

Datos fuente

Lectura de datos iniciales y su tratamiento inicial

```
# censo de población y vivienda de INEGI como archivo de texto
censo_2010A <- read.delim("datos/ITER_14TXT10.txt")
str(censo_2010A)
```

```
## 'data.frame':    11314 obs. of  201 variables:
## $ ENTIDAD      : int  14 14 14 14 14 14 14 14 14 14 ...
## $ NOM_ENT      : Factor w/ 1 level "Jalisco": 1 1 1 1 1 1 1 1 1 1 ...
## $ MUN          : int   0  0  0  1  1  1  1  1  1  1 ...
## $ NOM_MUN      : Factor w/ 126 levels "Acatic","Acatlán de Juárez",...: 105 105 105 1 1 1 1 1 1 1 ...
## $ LOC          : int   0 9998 9999  0  1  2  3  6  9 10 ...
## $ NOM_LOC      : Factor w/ 7067 levels "Abandono de Abajo",...: 6878 4240 4239 6879 14 70 105 1164 5898 1248 ...
## $ LONGITUD     : int   NA NA NA NA 1025419 1025331 1025103 1025849 1025619 1025217 ...
## $ LATITUD      : int   NA NA NA NA 204646 204737 204925 204344 204154 204734 ...
## $ ALTITUD      : int   NA NA NA NA 1693 1735 1794 1708 1630 1717 ...
## $ POBTOT       : int  7350682 16735 11116 21206 11890 107 10 36 215 27 ...
## $ POBMAS       : Factor w/ 919 levels "*","1","10","100",...: 518 17 709 21 706 679 1 265 36 155 ...
## $ POBFEM       : Factor w/ 923 levels "*","0","1","10",...: 528 759 675 32 732 673 1 213 29 113 ...
## $ P_0A2        : Factor w/ 440 levels "*","0","1","10",...: 273 390 363 74 393 421 1 3 61 152 ...
## $ P_0A2_M      : Factor w/ 340 levels "*","0","1","10",...: 101 196 171 286 192 194 1 2 282 3 ...
## $ P_0A2_F      : Factor w/ 339 levels "*","0","1","10",...: 103 188 175 280 196 232 1 3 261 3 ...
## $ P_3YMAS      : Factor w/ 1159 levels "*","0","10","100",...: 973 231 25 333 63 1151 1 642 347 467 ...
```

Proyecto 1 (preparar datos)

```
# tabla de menor tamaño para mayor comodidad
censo_2010A_sel <- censo_2010A[,c("MUN", "NOM_MUN", "LOC", "LONGITUD",
                                  "LATITUD", "POBTOT", "POBMAS", "POBFEM")]

# sustituir simbolos de * con NA
censo_2010A_sel[censo_2010A_sel == "*"] <- NA

# convertir el formato de columnas a numericas
censo_2010A_sel$POBFEM <- as.numeric(as.character(censo_2010A_sel$POBFEM))
censo_2010A_sel$POBMAS <- as.numeric(as.character(censo_2010A_sel$POBMAS))

# consultar tamaño de tabla
dim(censo_2010A_sel)
```

```
## [1] 11314      8
```

Proyecto 1 (eliminar registros innecesarios y clasificar localidades en menores y mayores)

```
# quitar los registros con el código de localidad 0 (corresponden a subtoales)
censo_2010A_sel <- censo_2010A_sel[censo_2010A_sel$LOC != 0
                                   & censo_2010A_sel$LOC != 9999
                                   & censo_2010A_sel$LOC != 9998,]

# consultar tamaño de tabla despues de limpieza
dim(censo_2010A_sel)
```

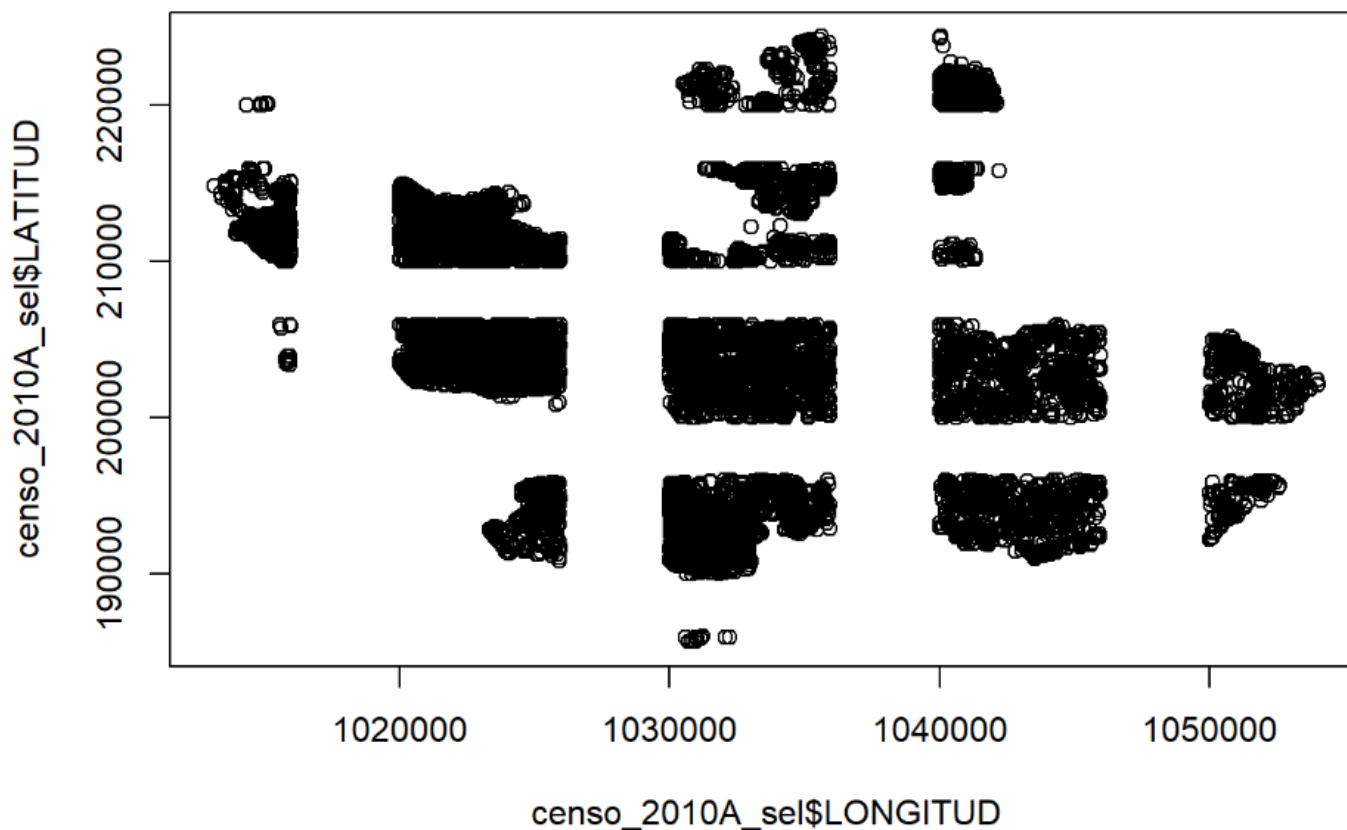
```
## [1] 10946      8
```

```
censo_2010A_sel$POBL_MENORES <- censo_2010A_sel$POBTOT < 50
```


Proyecto 1 (coordenadas, paso opcional)

Revisión de georeferenciación

```
# visualizar datos de latitud y longitud como estan en la tabla original  
plot(censo_2010A_sel$LONGITUD, censo_2010A_sel$LATITUD)
```



Proyecto 1 (coordenadas, paso opcional)

```
#a1 <- substr(censo_2010A_sel$LONGITUD,1,3)
#a2 <- substr(censo_2010A_sel$LONGITUD,4,5)
#a3 <- substr(censo_2010A_sel$LONGITUD,6,7)
#a1
#a2
#a3

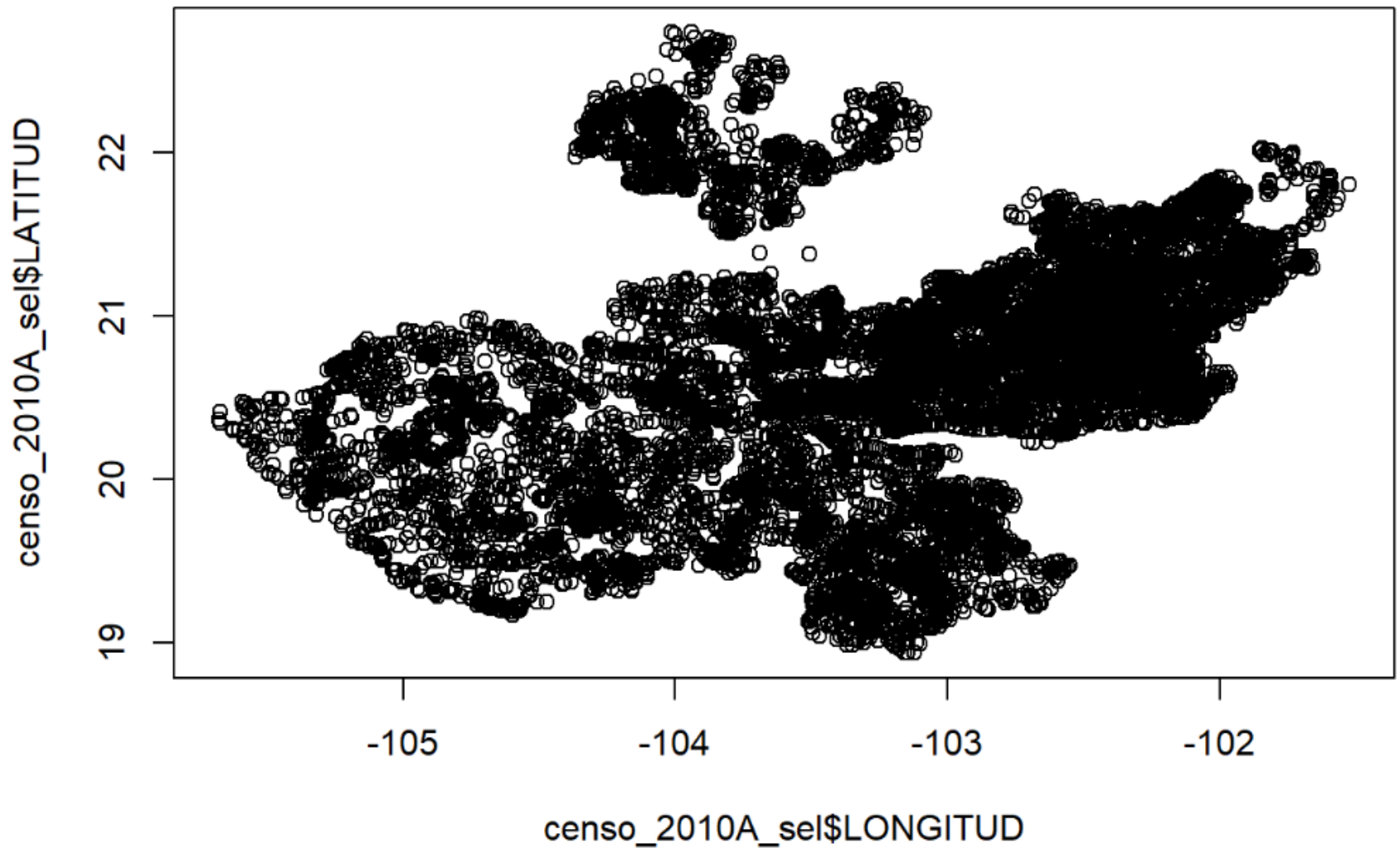
# transformar coordenadas al formato correcto
censo_2010A_sel$LONGITUD <- -1 * (as.numeric(substr(censo_2010A_sel$LONGITUD,1,3))
                                + as.numeric(substr(censo_2010A_sel$LONGITUD,4,5)) / 60
                                + as.numeric(substr(censo_2010A_sel$LONGITUD,6,7)) / 3600
                                )

censo_2010A_sel$LATITUD <- ( as.numeric(substr(censo_2010A_sel$LATITUD,1,2))
                             + as.numeric(substr(censo_2010A_sel$LATITUD,3,4)) / 60
                             + as.numeric(substr(censo_2010A_sel$LATITUD,5,6)) / 3600
                             )

#censo_2010A_sel$LONGITUD
#censo_2010A_sel$LATITUD

# visualizar datos despues de transformación de coordenadas
plot(censo_2010A_sel$LONGITUD, censo_2010A_sel$LATITUD)
```

Proyecto 1 (coordenadas, paso opcional)



Proyecto 1 (preparar datos)

Leer y preparar poligonos de municipios

```
# Leer municipios (Shapefile)
municipios <- readOGR("datos/Municipios.shp", encoding = "UTF-8")
```

```
## OGR data source with driver: ESRI Shapefile
## Source: "C:\Users\vshal\GD\UdeG_Docencia\CUCSH_Curso_R\sources\datos\Municipios.shp",
layer: "Municipios"
## with 2456 features
## It has 4 fields
```

```
# revisar estructura de tabla de atributos
str(municipios@data)
```

```
## 'data.frame':    2456 obs. of  4 variables:
## $ CVE_ENT      : Factor w/ 32 levels "01","02","03",...: 9 9 9 9 9 9 9 9 9 9 ...
## $ CVE_MUN      : Factor w/ 570 levels "001","002","003",...: 12 13 8 2 14 15 10 5 4 16
## ...
## $ NOM_MUN      : Factor w/ 2316 levels "Abalá","Abasolo",...: 2065 2241 816 185 206 456 7
3 610 450 945 ...
## $ CVE_MUNENT: int   9012 9013 9008 9002 9014 9015 9010 9005 9004 9016 ...
```

Proyecto 1 (preparar y revisar datos)

```
# selecciona solo municipios de Jalisco (entidad 14)  
municipios_jalisco <- municipios[municipios@data$CVE_ENT == '14',]  
municipios_jalisco@data$CVE_MUN <- as.numeric(municipios_jalisco@data$CVE_MUN)  
  
# revisar capa de municipios  
plot(municipios_jalisco)
```



Proyecto 1 (revisar datos)

```
summary(municipios_jalisco)
```

```
## Object of class SpatialPolygonsDataFrame
## Coordinates:
##           min      max
## x 2115863.6 2550361
## y  770594.3 1193103
## Is projected: TRUE
## proj4string :
## [+proj=lcc +lat_1=17.5 +lat_2=29.5 +lat_0=12 +lon_0=-102
## +x_0=2500000 +y_0=0 +ellps=GRS80 +units=m +no_defs]
## Data attributes:
```

##	CVE_ENT	CVE_MUN	NOM_MUN	CVE_MUNENT
## 14	:125	Min. : 1	Acatic : 1	Min. :14001
## 01	: 0	1st Qu.: 32	Acatlán de Juárez : 1	1st Qu.:14032
## 02	: 0	Median : 63	Ahualulco de Mercado: 1	Median :14063
## 03	: 0	Mean : 63	Amacueca : 1	Mean :14063
## 04	: 0	3rd Qu.: 94	Amatitán : 1	3rd Qu.:14094
## 05	: 0	Max. :125	Ameca : 1	Max. :14125
## (Other):	0		(Other) :119	

Proyecto 1 (resumir datos, paso opcional)

Análisis y resultados

Rvisar estaísticas generales de población, clasificar en localidades en menores (<50 habitantes) y mayores, calcular subtotales por municipio y por tipo de poblacion, calcular proporción de mujeres

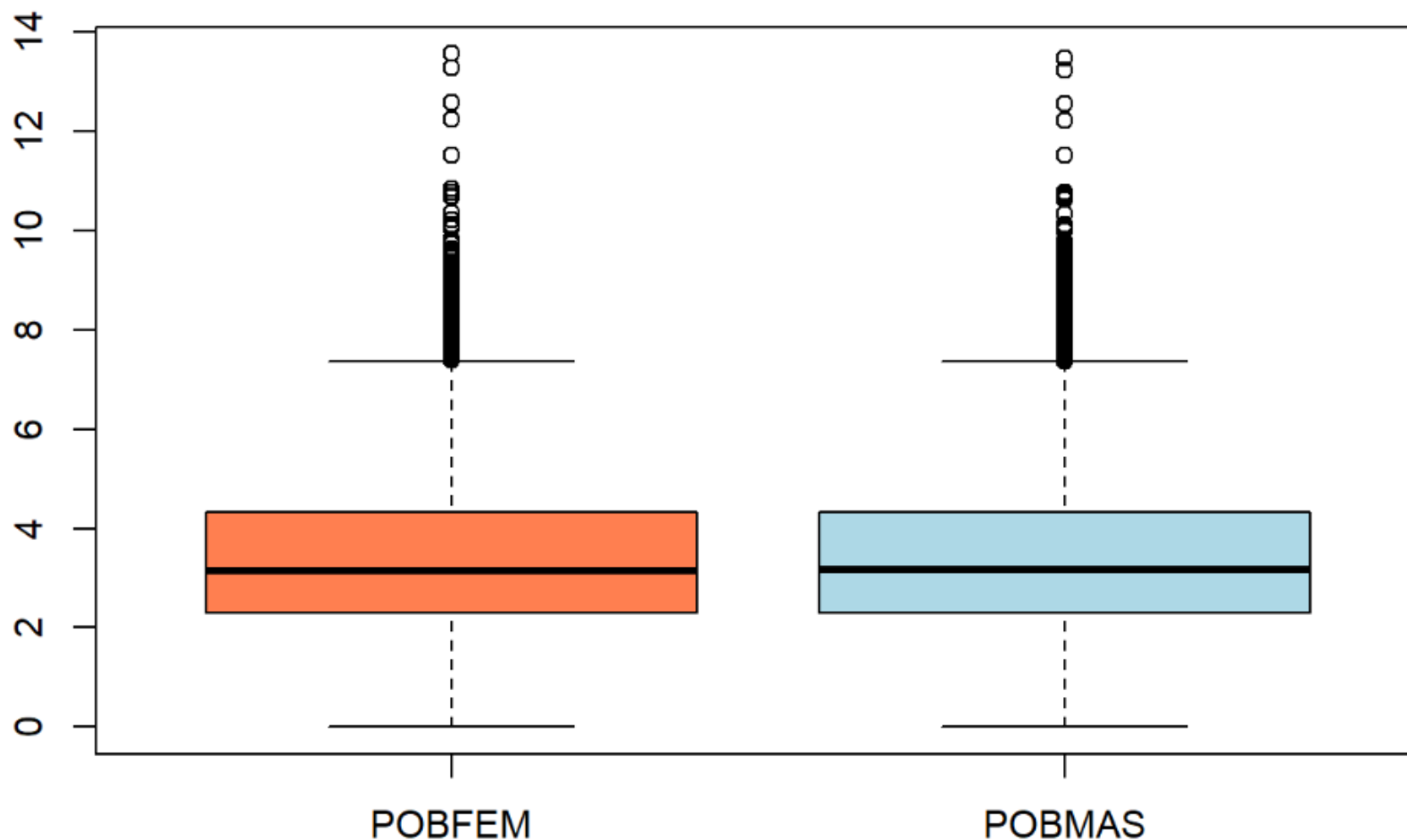
```
summary(censo_2010A_sel[,c("POBTOT", "POBFEM", "POBMAS")])
```

##	POBTOT		POBFEM		POBMAS
##	Min. :	1.0	Min. :	0.0	Min. : 1.0
##	1st Qu.:	5.0	1st Qu.:	10.0	1st Qu.: 10.0
##	Median :	14.0	Median :	23.0	Median : 24.0
##	Mean :	671.5	Mean :	598.9	Mean : 574.3
##	3rd Qu.:	60.0	3rd Qu.:	76.0	3rd Qu.: 76.0
##	Max. :	1495182.0	Max. :	777783.0	Max. : 717399.0
##			NA's :	4704	NA's : 4704

Adicionalmente se puede hacer graficas de POBFEM y POBMAS, calcular proporción de mujeres/hombres por localidad y visualizar distribución general de la proporción mencionada.

```
boxplot(log(censo_2010A_sel[,c("POBFEM","POBMAS")] ), col = c("coral","lightblue"))
```

```
## Warning in bplt(at[i], wid = width[i], stats = z$stats[, i], out =  
## z$out[z$group == : Outlier (-Inf) in boxplot 1 is not drawn
```



Proyecto 1 (calcular subtotales)

```
subtotales_municipios <- aggregate(cbind(POBFEM, POBMAS) ~  
                                   MUN + NOM_MUN + POBL_MENORES, data = censo_2010A_sel, sum)  
  
subtotales_municipios$FEM_PROP <- subtotales_municipios$POBFEM / (subtotales_municipios$POBFEM +  
subtotales_municipios$POBMAS)  
subtotales_municipios
```

##	MUN	NOM_MUN	POBL_MENORES	POBFEM	POBMAS	FEM_PROP
## 1	1	Acatic	FALSE	10245	9760	0.5121220
## 2	2	Acatlán de Juárez	FALSE	10899	10633	0.5061769
## 3	3	Ahualulco de Mercado	FALSE	10967	10598	0.5085555
## 4	4	Amacueca	FALSE	2670	2508	0.5156431
## 5	5	Amatitán	FALSE	7293	7187	0.5036602
## 6	6	Ameca	FALSE	29131	27804	0.5116536
## 7	8	Arandas	FALSE	35719	33231	0.5180421
## 8	10	Atemajac de Brizuela	FALSE	3277	3213	0.5049307
## 9	11	Atengo	FALSE	2553	2733	0.4829739

Proyecto 1 (reagrupar tablas)

Reacomodar y vincular datos en tablas

```
# generar dos tablas intermedias para poblaciones mayores y menores
subtotales_municipios_pobl_mayores <- subtotales_municipios[subtotales_municipios$POBL_MENORES == FALSE,
                                                             c("MUN", "POBFEM", "POBMAS", "FEM_PROP")]
subtotales_municipios_pobl_menores <- subtotales_municipios[subtotales_municipios$POBL_MENORES == TRUE,
                                                             c("MUN", "POBFEM", "POBMAS", "FEM_PROP")]

# realizar join de las dos tablas
subtotales_municipios_pobl <- merge(subtotales_municipios_pobl_mayores,
                                   subtotales_municipios_pobl_menores,
                                   by = "MUN", all = TRUE)
names(subtotales_municipios_pobl) <- c("MUN", "POBFEM_MAYORES", "POBMAS_MAYORES", "FEM_PROP_MAYORES",
                                       "POBFEM_MENORES", "POBMAS_MENORES", "FEM_PROP_MENORES")

# realizar JOIN de la tabla vinculada con la capa de poligonos de municipios
municipios_jalisco <- merge(municipios_jalisco,
                           subtotales_municipios_pobl,
                           by.x = "CVE_MUN", by.y = "MUN", all = TRUE)

head(municipios_jalisco@data)
```

Proyecto 1 (revisar tabla con resultados)

```
# revisar la tabla combinada
```

```
head(subtotales_municipios_pobl)
```

##	MUN	POBFEM_MAYORES	POBMAS_MAYORES	FEM_PROP_MAYORES	POBFEM_MENORES
## 1	1	10245	9760	0.5121220	532
## 2	2	10899	10633	0.5061769	69
## 3	3	10967	10598	0.5085555	43
## 4	4	2670	2508	0.5156431	127
## 5	5	7293	7187	0.5036602	56
## 6	6	29131	27804	0.5116536	134

##	POBMAS_MENORES	FEM_PROP_MENORES
## 1	509	0.5110471
## 2	76	0.4758621
## 3	48	0.4725275
## 4	117	0.5204918
## 5	55	0.5045045
## 6	149	0.4734982

Proyecto 1

```
head(municipios_jalisco@data)
```

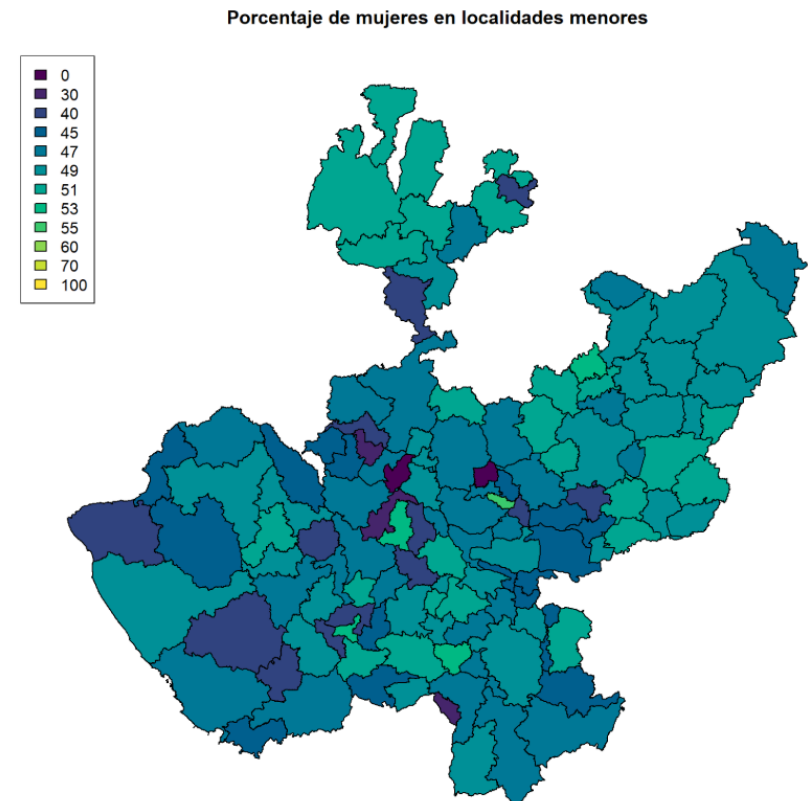
```
##      CVE_MUN CVE_ENT      NOM_MUN CVE_MUNENT POBFEM_MAYORES
## 67      67     14      Puerto Vallarta     14067      126754
## 43      43     14      La Huerta     14043      11167
## 81      81     14 Santa María de los Ángeles     14081      1927
## 41      41     14      Huejúcar     14041      2957
## 42      42     14      Huejuquilla el Alto     14042      4262
## 115     115     14      Villa Guerrero     14115      2572
##      POBMAS_MAYORES FEM_PROP_MAYORES POBFEM_MENORES POBMAS_MENORES
## 67      128174      0.4972149      254      305
## 43      11357      0.4957823      314      350
## 81      1706      0.5304156      36      47
## 41      2702      0.5225305      205      188
## 42      3966      0.5179874      239      221
## 115     2515      0.5056025      220      196
##      FEM_PROP_MENORES
## 67      0.4543828
## 43      0.4728916
## 81      0.4337349
## 41      0.5216285
## 42      0.5195652
## 115     0.5288462
```


Presentación de resultados

[illegible]

Proyecto 1 (generar mapa)

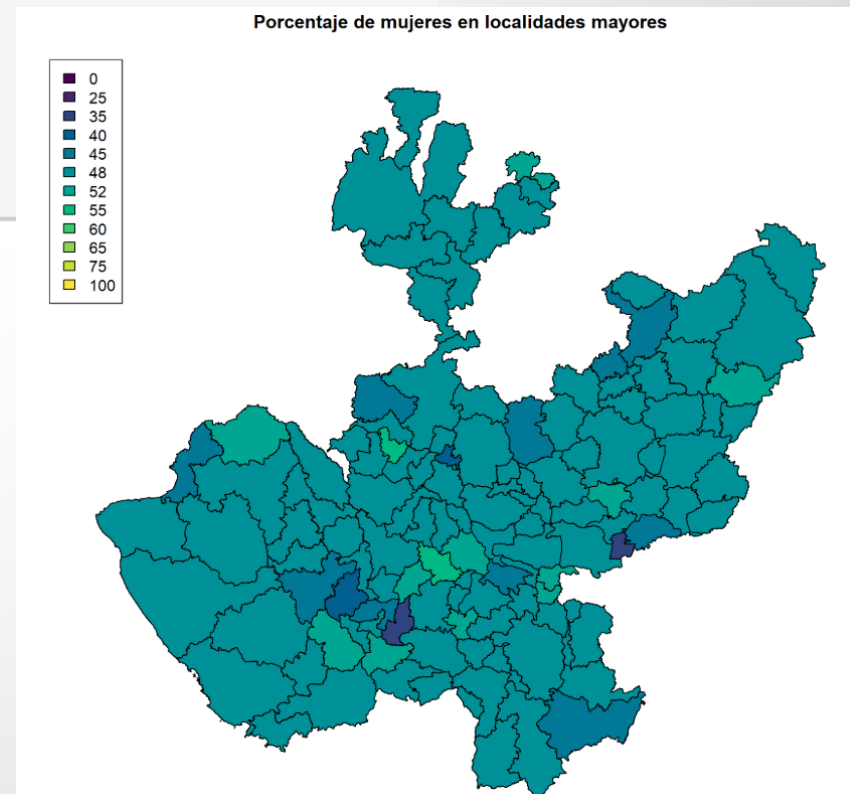
```
# visualizar mapas  
plot(municipios_jalisco, col = municipios_jalisco@data$Col1,  
     main = "Porcentaje de mujeres en localidades menores")  
  
legend("topleft", fill = hcl.colors(12),  
      #col = municipios_jalisco@data$Col1,  
      legend = rangos * 100)
```



Proyecto 1 (generar mapa)

```
plot(municipios_jalisco, col = municipios_jalisco@data$Col2,  
     main = "Porcentaje de mujeres en localidades mayores")
```

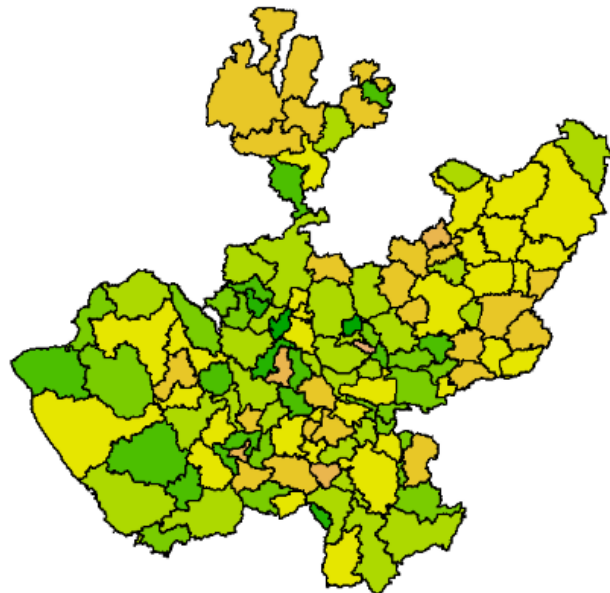
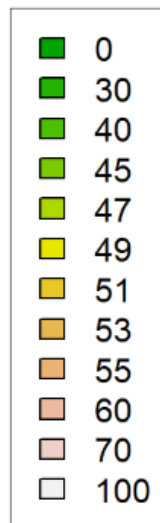
```
legend("topleft", fill = hcl.colors(12),  
      #col = municipios_jalisco@data$Col1,  
      legend = rangos * 100)
```



Alternativa de gradiente de colores con *terrain.colors()*

```
municipios_jalisco@data$Col1 <- terrain.colors(12)[as.numeric(  
  cut(municipios_jalisco@data$FEM_PROP_MENORES, rangos))]  
# visualizar mapas  
plot(municipios_jalisco, col = municipios_jalisco@data$Col1,  
     main = "Porcentaje de mujeres en localidades menores")  
  
legend("topleft", fill = terrain.colors(12),  
      legend = rangos * 100)
```

Porcentaje de mujeres en localidades menores

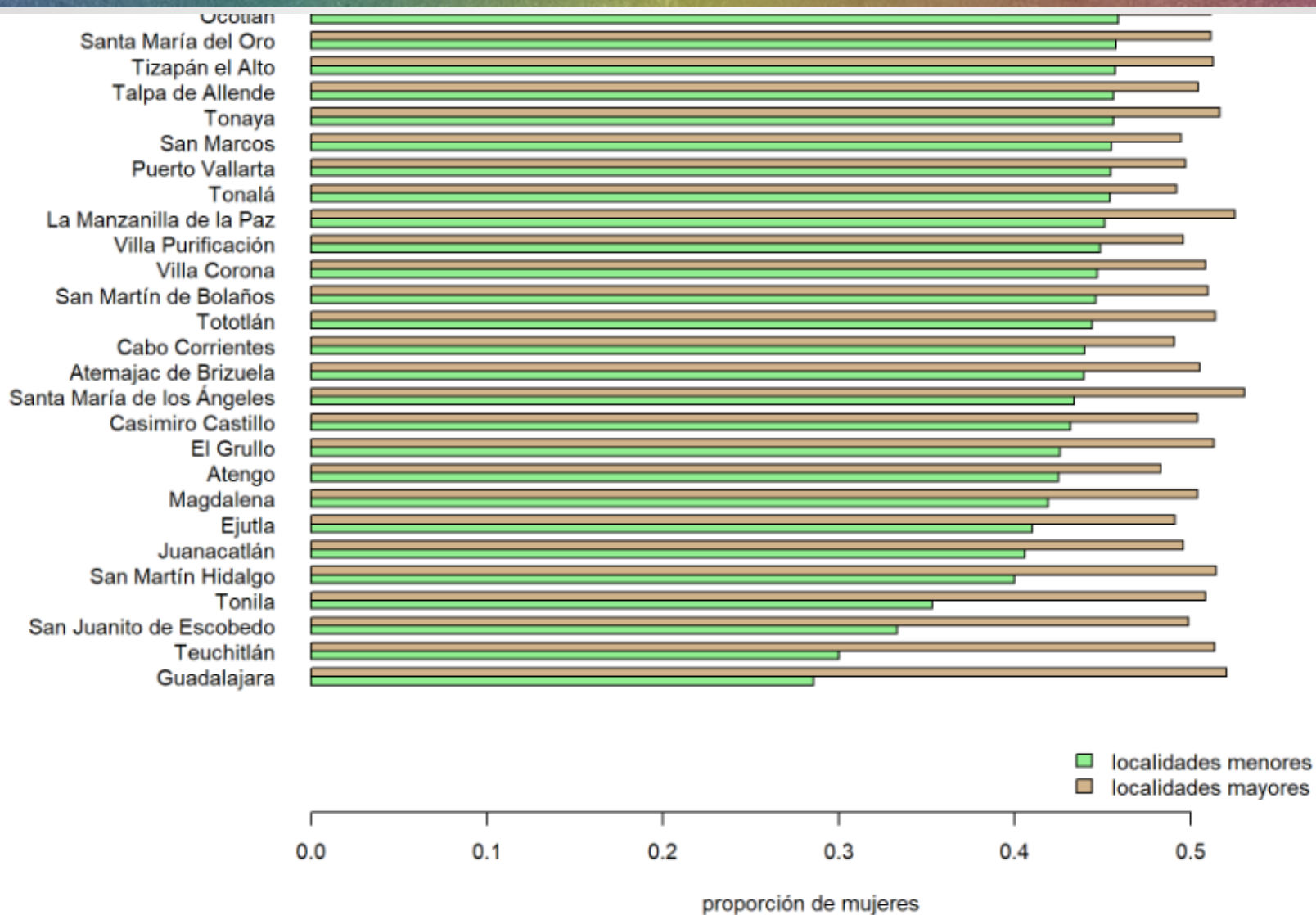


Proyecto 1 (generar gráfica, paso opcional)

```
par(mar=c(5,15,1,1), cex = 0.9)
# barplot
barplot(t(as.matrix(
  municipios_jalisco@data[
    order(municipios_jalisco@data$FEM_PROP_MENORES),
    c("FEM_PROP_MENORES", "FEM_PROP_MAYORES")
  ]
)),
  names.arg = municipios_jalisco@data[order(municipios_jalisco@data$FEM_PR
OP_MENORES), "NOM_MUN"],
  main = "Proporción mujeres/hombres por municipio de Jalisco",
  xlab = "proporción de mujeres",
  col = c("lightgreen", "tan"),
  beside = TRUE, horiz = TRUE, las = 1)

legend("bottomright", fill = c("lightgreen", "tan"), bty = "n",
  legend = c("localidades menores", "localidades mayores"))
```

Proyecto 1 (generar gráfica, paso opcional)





Gracias

vshalisko@gmail.com
Viacheslav Shalisko