

Instituto Tecnológico Autónomo de México

Proyecto de Probabilidad

Estudio probabilístico de la votación promedio por sección electoral del PAN en la CDMX

Elaborado por:

Alberto Reyes Briseño

Profesor: Mtro. Rodrigo Zepeda Tello

Proyecto Probabilidad

En el presente trabajo estudiamos la votación seccional promedio en general y por distrito, además, estudiamos impactos de las variables de edad, sexo e índice de escolaridad para determinar relaciones y probabilidades. ¿Cuántos votos en promedio por sección recibe el PAN en el agregado y en cada distrito uninominal en la Ciudad de México? ¿cuál es la probabilidad de que el PAN reciba X votos por sección en el agregado? ¿cuál es la relación de probabilidad que hay entre un cierto número de votación y el género? ¿cuál es la relación de probabilidad que hay entre el índice de escolaridad y la votación por sección? ¿en qué distritos es más probable que un público joven vote por el PAN?

En primer lugar, empezamos a identificar los distritos en el mapa que obtuvimos en el portal del INE.

Descripción de Bases de Datos utilizadas:

1. Históricos electorales de diputaciones de mayoría relativa desde el 2000 (INE)
<https://siceen21.ine.mx/home>
2. Corte del Padrón de Electores Nacional 2023 (INE)
<https://ine.mx/transparencia/datos-abiertos/#/archivo/datos-por-rangos-de-edad-entidad-de-origen-y-sexo-del-padron-electoral-y-lista-nominal-2023>
3. Índice de grado escolar por sección electoral en la Ciudad de México 2020 (INEGI)
http://gaia.inegi.org.mx/mdm6/?v=bGF0QjIzLjMyMDA4LGxvbjotMTAxLjUwMDAwLHo6MSxsOmNnZW9lbGVjdG9yYWxlcw==&layers=cgeoelectorales&fbclid=IwAR3_9hBLQGQw8rzaswU-Dw2SRfoIXiIBU_UbG2BOWUPQ4GpRwvMBMmMNHkI

Iniciación de bases de datos en R:

```
1 e2000 <- read_csv("2000_SEE_DIP_FED_MR_NAL_SEC.csv")
2 View(e2000)
3
4 e2006 <- read_csv("2006_SEE_DIP_FED_MR_NAL_SEC.csv")
5 View(e2006)
6
7 e2012 <- read_csv("2012_SEE_DIP_FED_MR_NAL_SEC.csv")
8 View(e2012)
9
10 e2018 <- read_csv("2018_SEE_DIP_FED_MR_NAL_SEC.csv")
11 View(e2000)
```

```

12
13 cat_secciones <- read_excel("9_Catalogo de Secciones con Distrito Local.xls")
14 View(cat_secciones)
15
16 escolaridad <- read_excel("Detalle de Estratos (1).xlsx",
17                           col_types = c("numeric", "text", "numeric",
18                                           "numeric"))
19 View(escolaridad)
20 edades_sexo <- read_excel("DatosAbiertos-derfe-pdln_edms_re_20230428.xlsx")
21 View(edades_sexo)
22 ##Crear frames exclusivamente de la Ciudad de México
23
24
25 e2018cm <- e2018 |>
26   filter(NOMBRE_ESTADO == "CIUDAD DE MÉXICO")
27
28
29
30 e2012cm <- e2012 |>
31   filter(NOMBRE_ESTADO == "DISTRITO FEDERAL")
32
33
34
35 e2006cm <- e2006 |>
36   filter(NOMBRE_ESTADO == "DISTRITO FEDERAL")
37
38 e2000cm <- e2000 |>
39   filter(NOMBRE_ESTADO == "DISTRITO FEDERAL")
40
41 edades_sexo <- edades_sexo |>
42   filter(NOMBRE_ENTIDAD == "CIUDAD DE MEXICO")

```

Desarrollamos la principal matriz que resulta del siguiente código:

```

1 e2000 <- ##SACAR PROMEDIO POR SECCIONES
2 ##LA TABLA BASE ES cat_secciones
3 base <- cat_secciones
4
5
6 base <- base |>
7   left_join(e2000cm |>
8             select(SECCION, AC), by = c("seccion" = "SECCION")) |>
9   rename(PAN2000 = AC)
10 base <- base |>
11   left_join(e2006cm |>
12             select(SECCION, PAN), by = c("seccion" = "SECCION")) |>
13   rename(PAN2006 = PAN)
14 base <- base |>
15   left_join(e2012cm |>
16             select(SECCION, PAN), by = c("seccion" = "SECCION")) |>
17   rename(PAN2012 = PAN)

```

```

18 base <- base |>
19 left_join(e2018cm |>
20           select(SECCION, PAN), by = c("seccion" = "SECCION")) |>
21 rename(PAN2018 = PAN)
22
23 #Sacamos el promedio de las votaciones presidenciales
24 base <- base |>
25 mutate(VPANprom = round((PAN2000+PAN2006+PAN2012+PAN2018)/4))
26
27 base <- base |> rowwise() |>
28 mutate(Desv = sd(c(PAN2000, PAN2006, PAN2012, PAN2018)))
29
30 #Modificar edades para tener menores de 29 años
31
32 edades_sexo <- edades_sexo |>
33 mutate(Menores_29 =
34 PADRON_18_HOMBRES+PADRON_18_MUJERES+PADRON_19_HOMBRES+
35 PADRON_19_MUJERES+PADRON_20_24_HOMBRES+PADRON_20_24_MUJERES+
36 PADRON_25_29_HOMBRES+PADRON_25_29_MUJERES) |>
37
38 mutate(Hombres_m29 = PADRON_18_HOMBRES+PADRON_19_HOMBRES+
39 PADRON_20_24_HOMBRES+
40 PADRON_25_29_HOMBRES) |>
41
42 mutate(Mujeres_m29 = PADRON_18_MUJERES+PADRON_19_MUJERES+
43 PADRON_20_24_MUJERES+
44 PADRON_25_29_MUJERES)
45
46
47
48
49
50
51
52
53
54
55
56

```

Elaboramos la gráfica de densidad de votos promedio por sección para el PAN tomando el promedio de las elecciones de 2000, 2006, 2012 y 2018 por ser presidenciales.

```

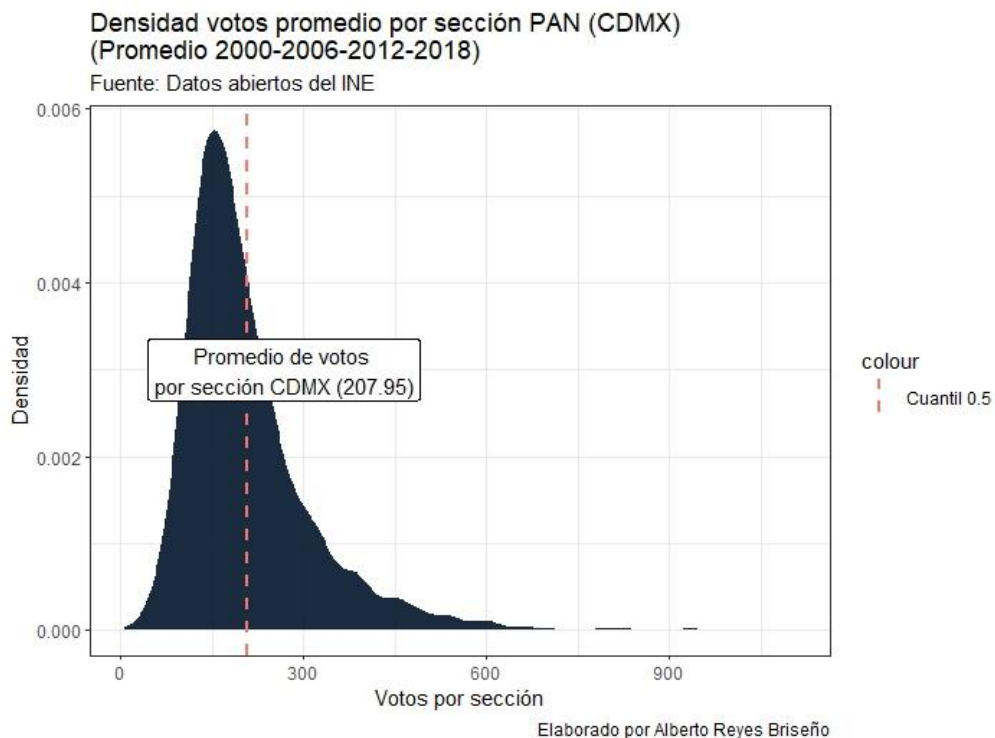
1 frecuencia_voto <- base |>
2   select(seccion, VPANprom)
3
4 frecuencia_voto |> mean(VPANprom, na.rm = T)
5
6 mean(frecuencia_voto, VPANprom, na.rm = TRUE)

```

```

1 ggplot(frecuencia_voto) +
2   geom_density(aes(x = VPANprom), fill = "#02152b", alpha = 0.9, color = NA) +
3   geom_vline(aes(xintercept = mean(frecuencia_voto$VPANprom, na.rm = T),
4     color = "Cuantil 0.5"),
5     linetype = "dashed",
6     size = 1) +
7   annotate("label", x = mean(frecuencia_voto$VPANprom, na.rm = T) + 60, y =
8     .003,
9     label = "Promedio de votos
10  por sección CDMX (206)", angle = 90) +
11   labs(
12     x = "Votos por sección",
13     y = "Densidad",
14     title = "Densidad votos promedio por sección PAN (CDMX)
15  (Promedio 2000-2006-2012-2018)",
16     subtitle = "Fuente: Datos abiertos del INE",
17     caption = "Elaboración Propia"
18   ) +
19   theme_bw()

```

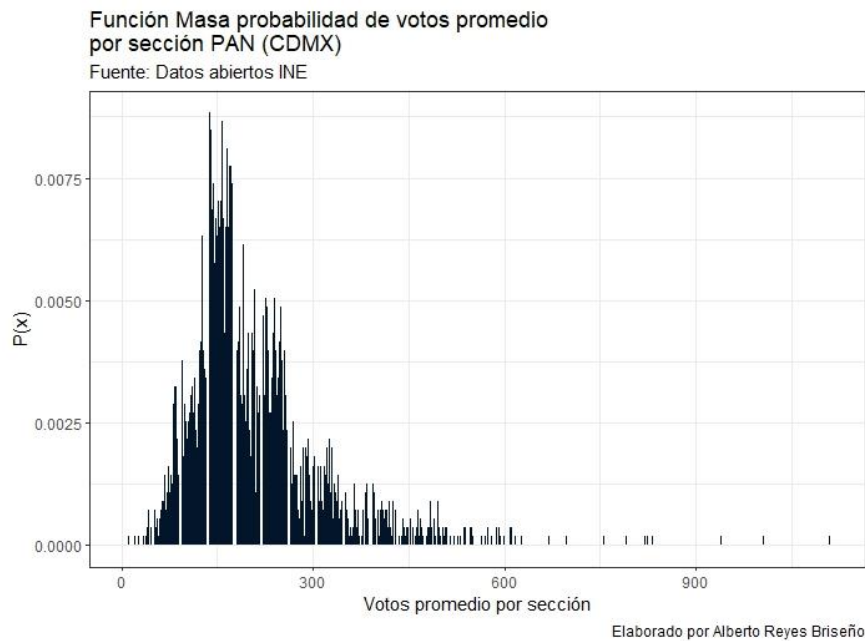


Creamos el data frame de la función de masa de probabilidad con el siguiente código y creamos la gráfica:

```

1 #Calculo de probabilidad voto por sección promedio en CDMX
2 prob_votos_seccion <- base |>
3   group_by(VPANprom) |>
4   tally()
5 ggplot(prob_votos_seccion) +
6   geom_col(aes(x = VPANprom, y = n))
7
8 prob_votos_seccion <- prob_votos_seccion |>
9   mutate(Probabilidad = n / sum(n))
10 ##Gráfica de masa de probabilidad
11 ggplot(prob_votos_seccion) +
12   geom_col(aes(x = VPANprom, y = Probabilidad), fill = "#02152b") +
13   labs(
14     x = "Votos promedio por sección",
15     y = "P(x)",
16     title = "Función Masa probabilidad de votos promedio
17 por sección PAN (CDMX)",
18     subtitle = "Fuente: Datos abiertos INE",
19     caption = "Elaboración propia"
20   ) +
21   theme_bw()

```



Con base en esto, podemos calcular la probabilidad acumulada donde X es una variable aleatoria que representa los votos promedio por sección (CDMX).

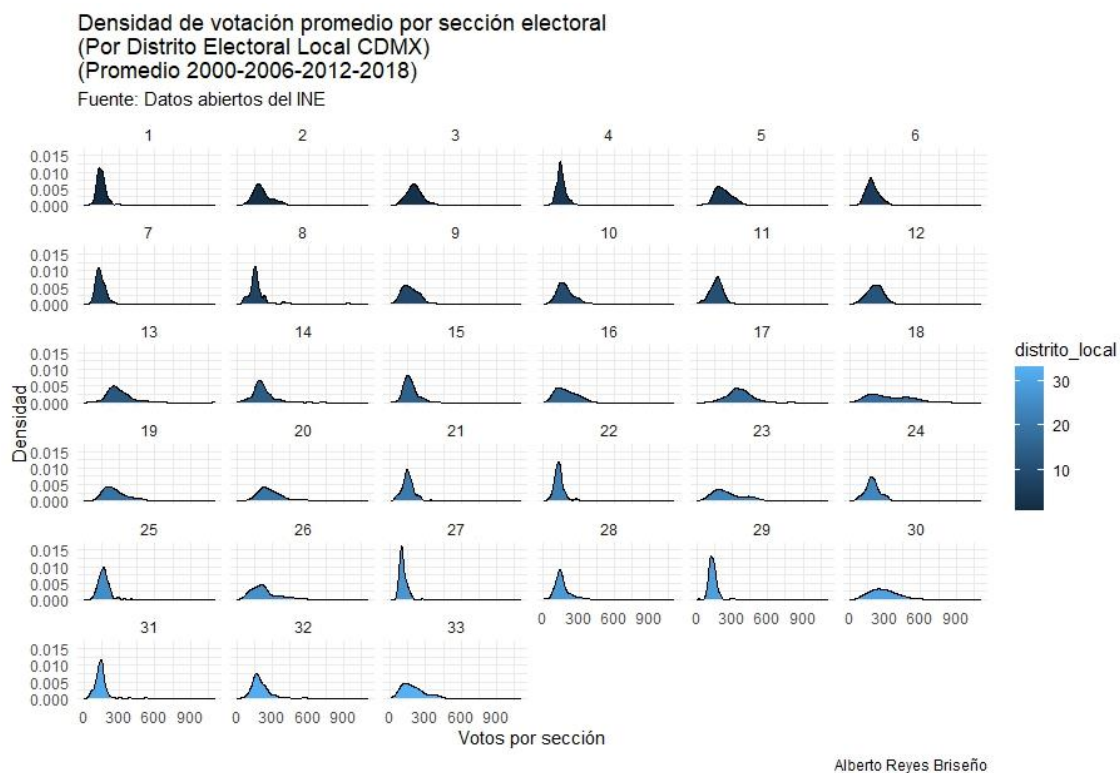
Por ejemplo **$P(X \geq 300)$**

```

1 #Calculo de probabilidad acumulada
2 #Ejemplo de P(X >= 300)
3
4 prob_votos_seccion |>
5   filter(VPANprom >= 300) |>
6   summarise(
7     "Probabilidad más de 160 votos" =
8       scales::percent(sum(Probabilidad))
9   )

```

De forma general, mostramos las funciones de densidad de cada distrito uninominal local de la Ciudad de México.



Sacamos las medidas estadísticas por distrito

```

1 summarise_dtos <- base |>
2   group_by(distrito_local) |>
3   summarise(
4     "Media" = mean(VPANprom, na.rm = T),
5     "SD" = sd(VPANprom, na.rm = T),
6     "Mediana" = median(VPANprom, na.rm = T),
7     "Min" = min(VPANprom, na.rm = T),
8     "Max" = max(VPANprom, na.rm = T)
9   )
10
11 mean_dto <- summarise_dtos |>

```

```

12 select(distrito_local, Media)
13 ##Hacer tabla de promedios por distrito
14 xlsx::write.xlsx(mean_dto, file = "mean_dto.xlsx", row.names = T)
15 mean_dto <- mean_dto |>
16 pivot_wider(names_from = distrito_local, values_from = Media)

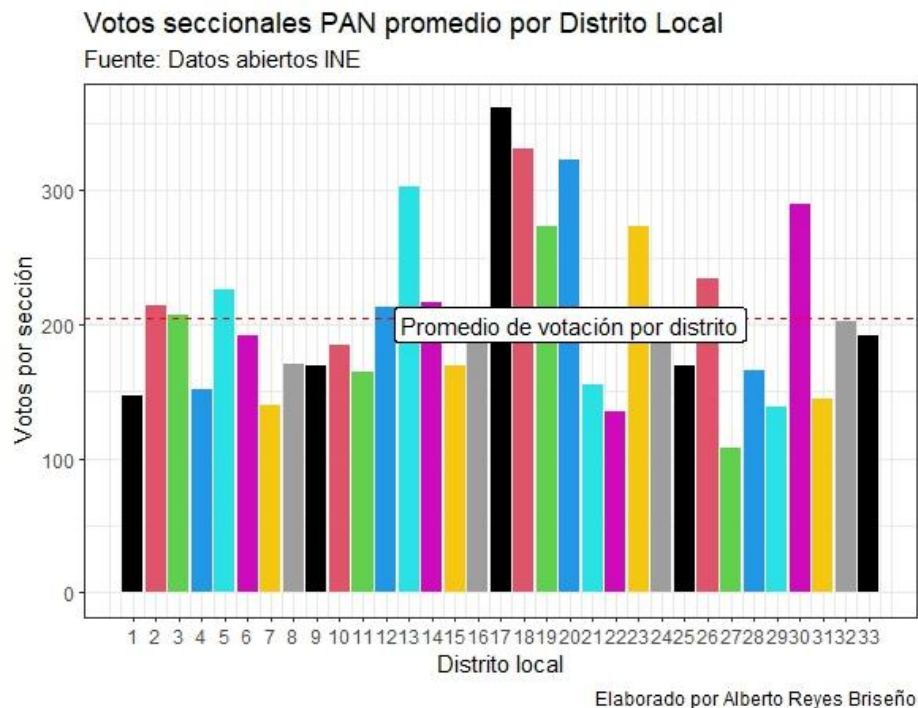
```

Con esta información, elaboramos un histograma que muestre la votación seccional promedio por distrito y su media. Podemos observar que los distritos más rentables para el PAN, son aquellos en donde hay una mayor votación promedio por sección.

```

##Histograma media
summarise_dtos |>
1  ggplot() +
2    geom_col(aes(x = distrito_local, y = Media), fill =
3  summarise_dtos$distrito_local) +
4    scale_fill_brewer(palette = "Set1")+
5    geom_hline(aes(yintercept = mean(Media, na.rm = T)),
6              linetype = "dashed",
7              color = "red") +
8    annotate("label", x = 20, y = 200,
9            label = "Promedio de votación por distrito", angle = 90) +
10   labs(
11     x = "Distrito local",
12     y = "Votos por sección",
13     title = "Votos seccionales PAN promedio por Distrito Local",
14     subtitle = "Fuente: Datos abiertos INE",
15     caption = "Elaboración propia"
16   ) +
17   scale_x_continuous("Distrito local", labels =
18 as.character(summarise_dtos$distrito_local), breaks =
19 summarise_dtos$distrito_local)+
  theme_bw()

```

¿Quiénes votan por el PAN en cada sección electoral? Usaremos el índice de escolaridad por sección electoral que el INEGI calculó en el 2020 para determinad las funciones masa probabilidad de la variable aleatoria X que representa los votos que obtiene el PAN en una sección electoral dado su grado de escolaridad. Para lo anterior, mostramos el catálogo que el INEGI provee para la interpretación del coeficiente de escolaridad.

Nivel de instrucción		Años acumulados (grado de escolaridad)
Sin Preparación		0
Primaria	1	1
	2	2
	3	3
	4	4
	5	5
	6	6
Secundaria	1	7
	2	8
	3	9
Bachillerato	1	10
	2	11
	3	12
Universidad (Profesional)		13 o mas

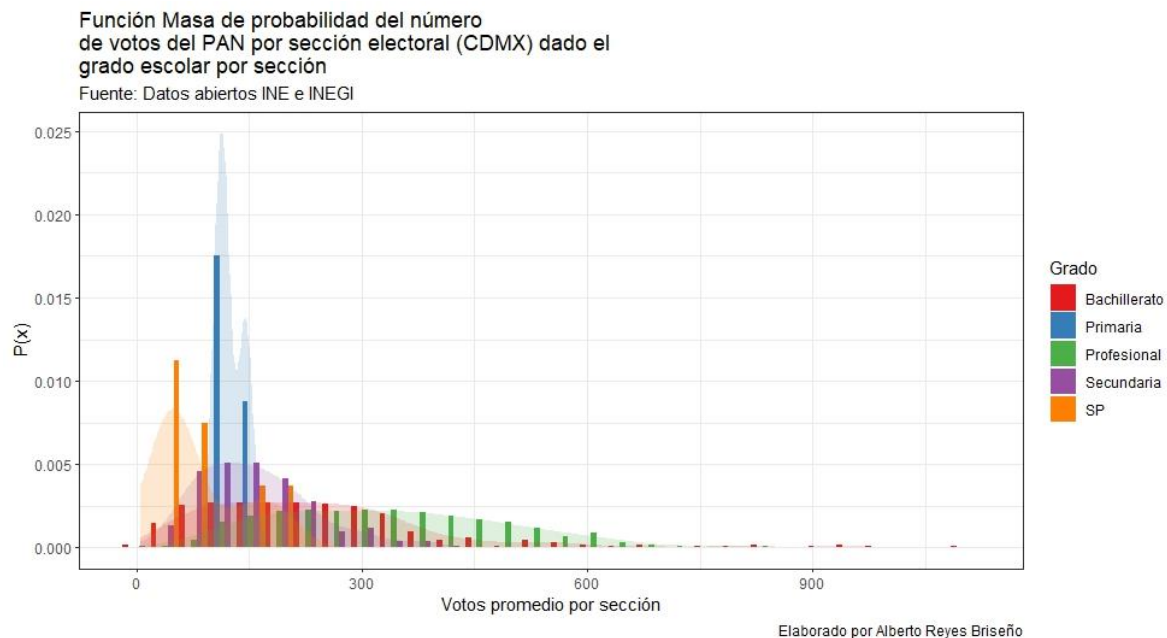
Procedemos a elaborar el data frame que nos sirve para tal propósito

```

## FUNCIONES MASA DE PROBABILIDAD POR GRADO DE ESCOLARIDAD
1 # Creamos el DataFrame
2 grado <- base |>
3   select(seccion, VPANprom, Escolaridad)
4
5 grado <- grado |>
6   mutate(
7     Grado = if_else(Escolaridad == 0, "SP",
8                     if_else(Escolaridad > 0 & Escolaridad < 7, "Primaria",
9                             if_else(Escolaridad >= 7 & Escolaridad < 10,
10 "Secundaria",
11                                     if_else(Escolaridad >= 10 & Escolaridad
12 < 13, "Bachillerato", "Profesional"))))
13   )
14
15 grado_masa <- grado |>
16   group_by(VPANprom, Grado) |>
17   tally() |>
18   rename(Freq = n)

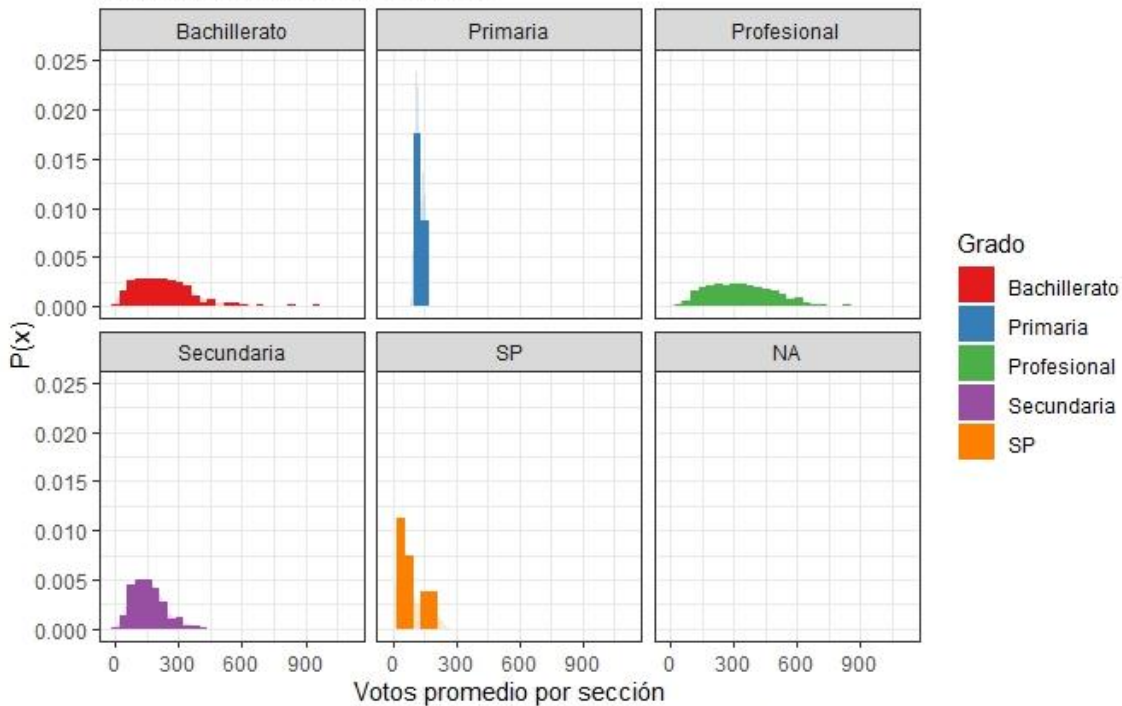
```

Con lo anterior, procedemos a elaborar las gráficas de las funciones de masa de probabilidad de cada grado escolar en donde X es una variable aleatoria que representa el número de votos promedio en la casilla, así como sus estadísticos.



Función Masa de probabilidad del número de votos del PAN por sección electoral (CDMX) dado el grado escolar por sección

Fuente: Datos abiertos INE e INEGI



Elaborado por Alberto Reyes Briseño

Grado	Media	SD	Mediana	Min	Max
Bachillerato	244.3934	166.5961	219.5	11	1099
Primaria	123.6667	18.90326	117	109	145
Profesional	340.2633	161.3635	326	35	1110
SP	83.57143	61.50842	58	28	192
Secundaria	157.9082	76.75563	148.5	6	401

De lo anterior concluimos las siguientes cosas:

1. Las secciones electorales con un coeficiente de grado escolar promedio por encima de 13.0 (Universidad o Profesión) concentran más probabilidad de registrar una votación alta para el PAN en promedio, sin embargo, su distribución nos dice que también donde hay poca votación es probable que se cuente con este grado de estudios. Donde el PAN registra más votación (propia, no relativa) es en las secciones con grados de estudios universitarios, empero, no significa que en el agregado esta población solo vote por el PAN (son casi igual de probables).
2. Las secciones electorales en donde el grado de escolaridad oscila entre sin preparación y la primaria, la probabilidad de votación del PAN se concentra entre

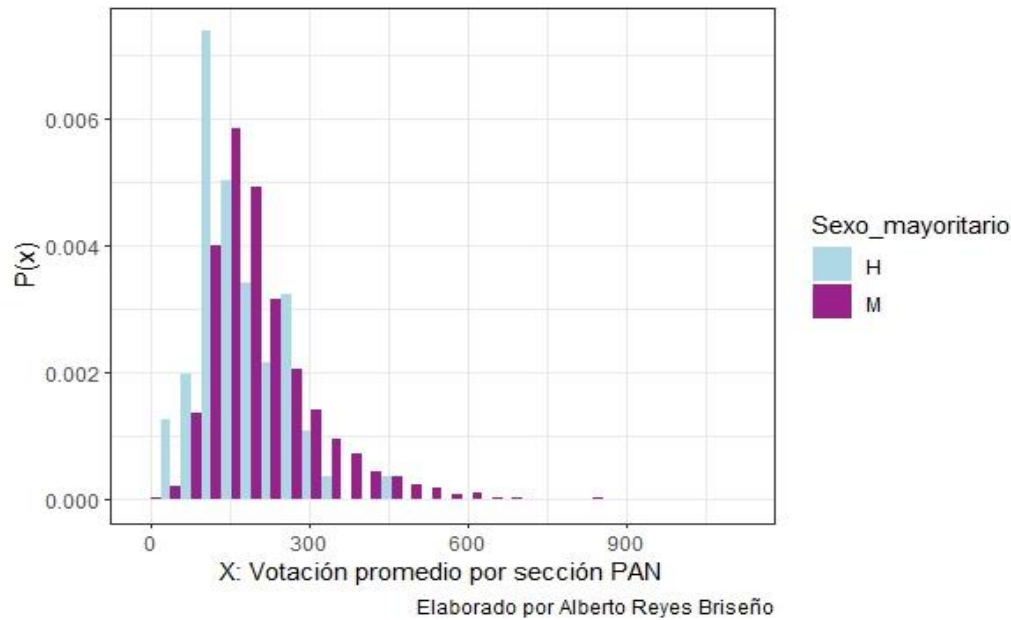
las más bajas, es decir, donde el PAN tiene pocos votos el grado de escolaridad promedio está entre los ya mencionados.

Ahora trabajaremos con la frecuencia de hombres y mujeres en cada sección electoral que el INE registró en su último corte de Padrón Nacional. La función de masa de probabilidad de los votos seccionales promedio dado que la sección tenga una mayoría absoluta ya sea de hombres o mujeres está dado por:

```
## FUNCIONES MASA DE PROBABILIDAD DONDE LA MAYORIA ABSOLUTA SEA DE UN SEXO
1 (50 + 1)
2
3 masa_sexo <- base
4
5
6 masa_sexo <- masa_sexo |>
7   mutate(Sexo_mayoritario = if_else(
8     PADRON_HOMBRES/PADRON_ELECTORAL >= .5, "H", "M"
9   ))
10
11 ##Numero de secciones que el hombre o la mujer es mayoría absoluta
12 masa_sexo |>
13   group_by(VPANprom, Sexo_mayoritario) |>
14   tally()
15
16 masa_sexo <- masa_sexo |>
17   select(VPANprom, Sexo_mayoritario)
18
19 ggplot(masa_sexo) +
20   geom_histogram(aes(x = VPANprom, y = after_stat(density), fill =
21     Sexo_mayoritario),
22     position = position_dodge()) +
23   labs(
24     x = "X: Votación promedio por sección PAN",
25     y = "P(x)",
26     title = "Función Masa de probabilidad de votos PAN por sección
27     dado el sexo mayoritario (50% + 1)",
28     subtitle = "Fuente: Datos abiertos del INE",
29     caption = "Elaboración propia"
30   ) +
31   scale_fill_manual(values = c("H" = "lightblue", "M" = "#9A208C")) +
32   theme_bw()
```

Función Masa de probabilidad de votos PAN por sección dado el sexo mayoritario (50% + 1)

Fuente: Datos abiertos del INE



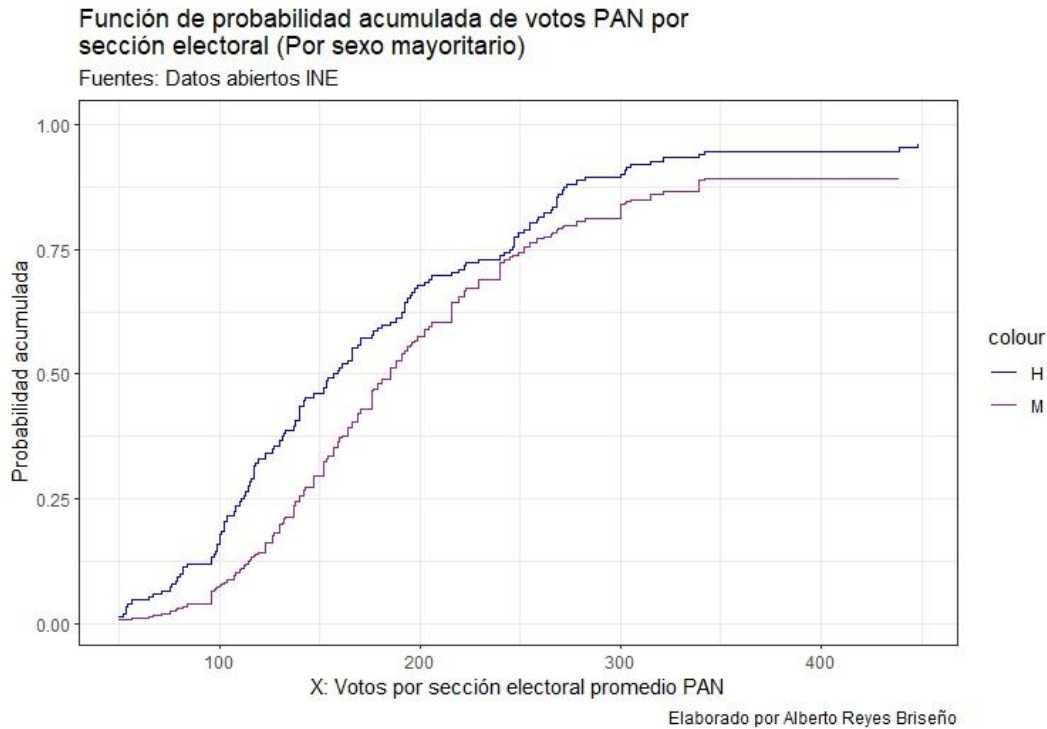
También elaboramos las funciones de probabilidad acumulada por sexo mayoritario para visualizar mejor el comportamiento.

```
1 #Elaboramos la gráfica de probabilidad acumulada para cada sexo mayoritario
2
3 mh <- masa_sexo |>
4   filter(Sexo_mayoritario == "H") |>
5   group_by(VPANprom, Sexo_mayoritario) |>
6   tally()
7 mh <- mh |>
8   mutate(prob = n / sum(mh$n))
9 cdf <- function(x, p){
10   cumsum(p)
11 }
12
13 Fa <- cdf(mh$VPANprom, mh$prob)
14 funh <- data.frame(mh$VPANprom, Fa)
15
16 ggplot(funh) +
17   geom_step(aes(x = mh.VPANprom, y = Fa), color = "blue") +
18   labs(
19     x = "X: Votos por sección electoral promedio PAN",
20     y = "Probabilidad acumulada",
21     title = "Función de probabilidad acumulada de votos PAN por
22 sección electoral ( 50 + 1 son Hombres)",
23     subtitle = "Fuentes: Datos abiertos INE",
24     caption = "Elaboración propia"
25   ) +
26   theme_bw()
```

```

27
28 #Ahora para mujer
29
30 mm <- masa_sexo |>
31   filter(Sexo_mayoritario == "M") |>
32   group_by(VPANprom, Sexo_mayoritario) |>
33   tally()
34 mm <- mm |>
35   mutate(prob = n / sum(mm$n))
36 cdf <- function(x, p){
37   cumsum(p)
38 }
39
40 Fa <- cdf(mm$VPANprom, mm$prob)
41 funm <- data.frame(mm$VPANprom, Fa)
42
43 ggplot(funm) +
44   geom_step(aes(x = mm.VPANprom, y = Fa), color = "#9A208C") +
45   labs(
46     x = "X: Votos por sección electoral promedio PAN",
47     y = "Probabilidad acumulada",
48     title = "Función de probabilidad acumulada de votos PAN por
49 sección electoral ( 50 + 1 son Mujeres)",
50     subtitle = "Fuentes: Datos abiertos INE",
51     caption = " Elaboración propia "
52   ) +
53   theme_bw()
54
55 acu <- funh |>
56   left_join(funm, by = c("mh.VPANprom" = "mm.VPANprom")) |>
57   rename("H" = Fa.x) |>
58   rename("M" = Fa.y)
59
60 ggplot(acu) +
61   geom_step(aes(x = mh.VPANprom, y = H, color = "H")) +
62   geom_step(aes(x = mh.VPANprom, y = M, color = "M")) +
63   labs(
64     x = "X: Votos por sección electoral promedio PAN",
65     y = "Probabilidad acumulada",
66     title = "Función de probabilidad acumulada de votos PAN por
67 sección electoral (Por sexo mayoritario)",
68     subtitle = "Fuentes: Datos abiertos INE",
69     caption = " Elaboración propia "
70   ) +
71   scale_color_manual(values = c("H" = "blue", "M" = "#9A208C")) +
72   theme_bw()
73

```



La conclusión que se desprende del análisis es que hay más probabilidad de que el PAN tenga una votación alta en aquellas secciones donde más del 50% de la población es mujer.

Para analizar la relación entre el voto promedio por sección y el porcentaje de menores de 29 por sección electoral elaboramos el estadístico de la densidad conjunta de:

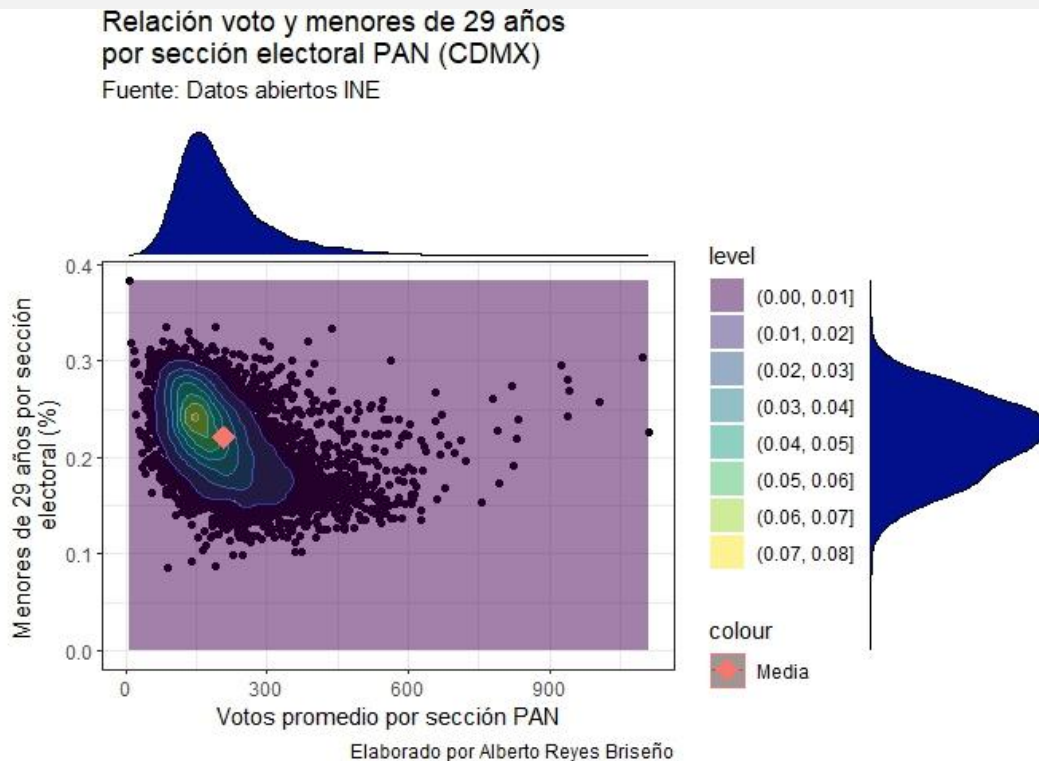
- Votos promedio por sección electoral; y
- % de menores de 29 años en cada sección electoral.

```
1 masa_con <- base |>
2   select(VPANprom, Menores_29, PADRON_ELECTORAL)
3
4 masa_con <- masa_con |>
5   mutate(Pormen = Menores_29 / PADRON_ELECTORAL)
6
7 med_con <- masa_con |>
8   summarise("VPANprom" = mean(masa_con$VPANprom, na.rm = T),
9             "Menores_29" = mean(masa_con$Menores_29, na.rm = T),
10            "Pormen" = mean(masa_con$Pormen, na.rm = T))
11
12
13 mean(masa_con$VPANprom, na.rm = T)
14
15 ## gráfica con marginales (votos por sección / % menores de 29)
16 p <- ggplot(masa_con, aes(VPANprom, Pormen)) +
17   geom_point() +
18   geom_density2d() +
19   geom_density2d_filled(alpha = 0.5) +
```

```

20 geom_point(aes(color = "Media"), shape = 18, size = 5, data =
21 med_con) +
22 labs(
23   x = "Votos promedio por sección PAN",
24   y = "Menores de 29 años por sección
25 electoral (%)",
26   title = "Relación voto y menores de 29 años
27 por sección electoral PAN (CDMX)",
28   subtitle = "Fuente: Datos abiertos INE",
29   caption = "Elaborado por Alberto Reyes Briseño"
30 ) +
31 theme_bw()
32
33 cor(masa_con$VPANprom, masa_con$Pormen, use = "complete.obs")
34
35
36
37
38

```



De lo anterior concluimos que, a menor porcentaje de menores de 29 años en la sección electoral, la probabilidad de tener una votación amplia disminuye.

Correlación

-0.4401755

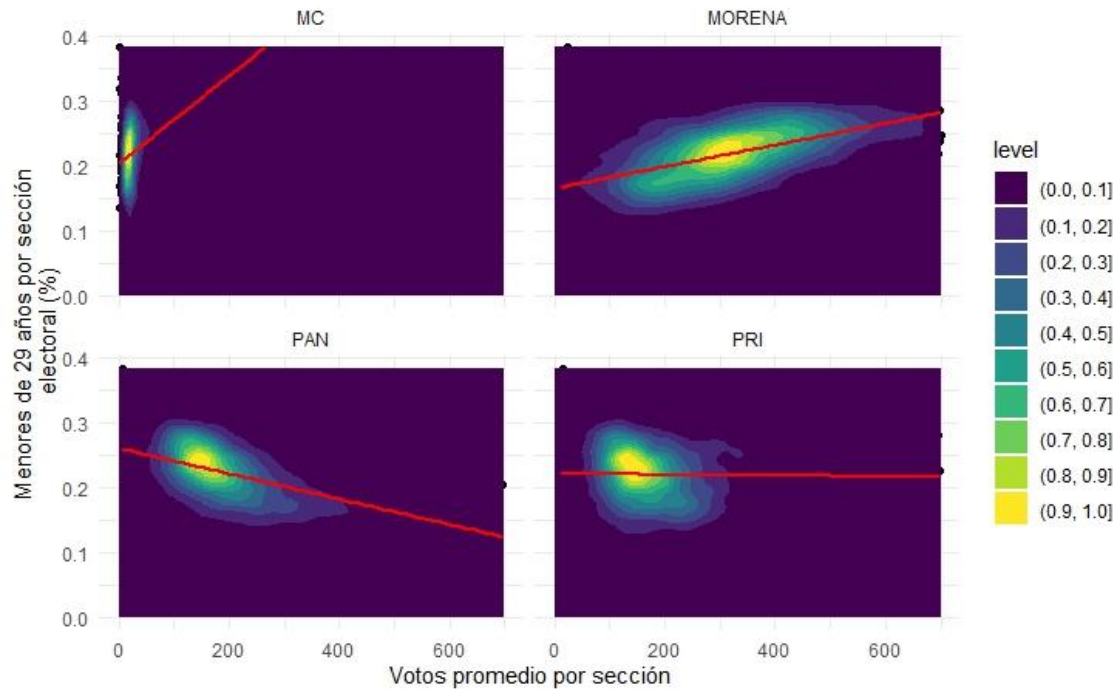
Por último, vamos a comparar las densidades conjuntas del PAN, MORENA y MC (votos por sección / % menores de 29 años)

```
1
2
3
4
5
6
7
8 ##Hacemos la comparación de conjuntas PAN-MORENA-MC (voto por
9 sección / % menores de
10 #29)
11
12 base_comp <- base |>
13   select(seccion, VPANprom, Menores_29, PADRON_ELECTORAL)
14
15 base_comp <- base_comp |>
16   left_join(e2018cm, by = c("seccion" = "SECCION")) |>
17   select(seccion, VPANprom, Menores_29, MORENA, , MC, Menores_29,
18 PADRON_ELECTORAL) |>
19   rename(MORENA18 = MORENA, MC18 = MC) |>
20   left_join(e2021cm, by = c("seccion" = "SECCION")) |>
21   select(seccion, VPANprom, Menores_29, MORENA18, MC18, MORENA, MC,
22 Menores_29, PADRON_ELECTORAL) |>
23   rename(MORENA21 = MORENA, MC21 = MC)
24
25 base_comp <- base_comp |>
26   mutate(VMORprom = round((MORENA18 + MORENA21)/2),
27           MC = round(MC18 + MC21) / 2 ) |>
28   select(seccion, VPANprom, VMORprom, Menores_29, PADRON_ELECTORAL,
29 MC) |>
30   mutate(promEd = Menores_29/PADRON_ELECTORAL) |>
31   rename(PAN = VPANprom, MORENA = VMORprom)
32
33 base_comp <- base_comp |>
34   pivot_longer(cols=c("PAN", "MORENA", "MC"),
35                 names_to = "partido",
36                 values_to = "votacion")
37
38
39 #####SUMMARY
40 sum_parti <- base_comp |>
41   group_by(partido) |>
42   summarise("Correlación" = cor(votacion, promEd, use =
43 "complete.obs"))
44
45
46
47
48
49
50
```

51
52
53

Relación voto por partido y % menores de 29 años por sección electoral (CDMX)

Fuente: Datos abiertos INE



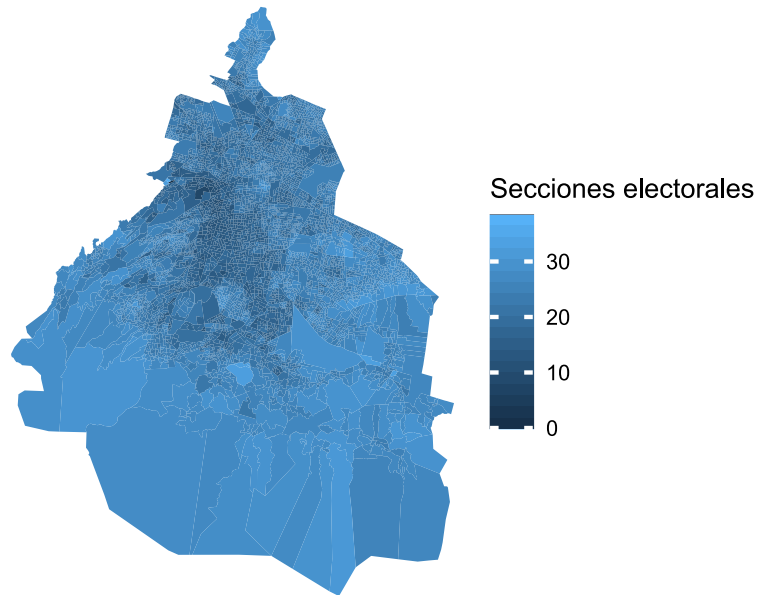
Elaborado por Alberto Reyes Briseño

Partido	Correlación
MC	0.309767654
MORENA	0.523412289
PAN	-0.440175513
PRI	0.0814

Con lo anterior concluimos que, mientras MORENA concentra más probabilidad de votación en secciones donde hay más jóvenes, el PAN tiene una relación inversa, es decir, en las secciones electorales donde hay mayor porcentaje de jóvenes su intención de voto es menor, y por ende, su probabilidad igual.

Porcentaje de jóvenes por sección electoral en la CDMX

Fuente: Datos abiertos INE

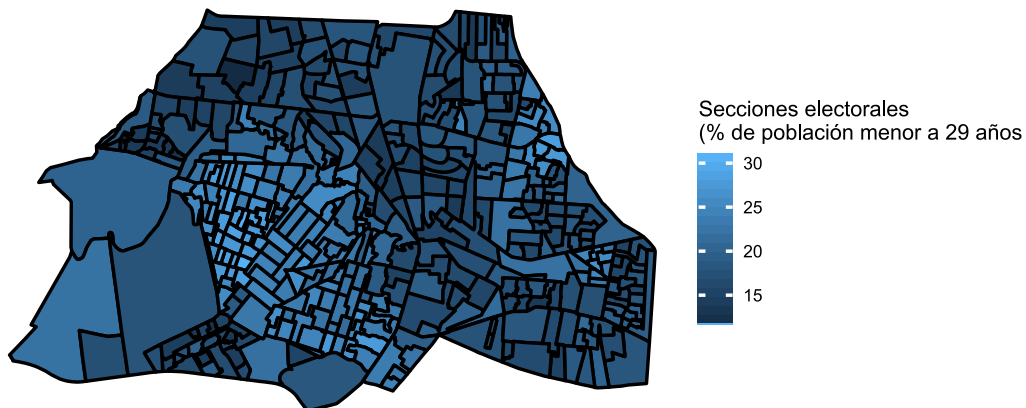


Elaborado por Alberto Reyes Briseño

Para más detalle, tomaremos a la alcaldía Coyoacán como referencia para aterrizar el comportamiento de la edad.

Porcentaje de jóvenes por sección electoral en Coyoacán

Fuente: Datos abiertos INE

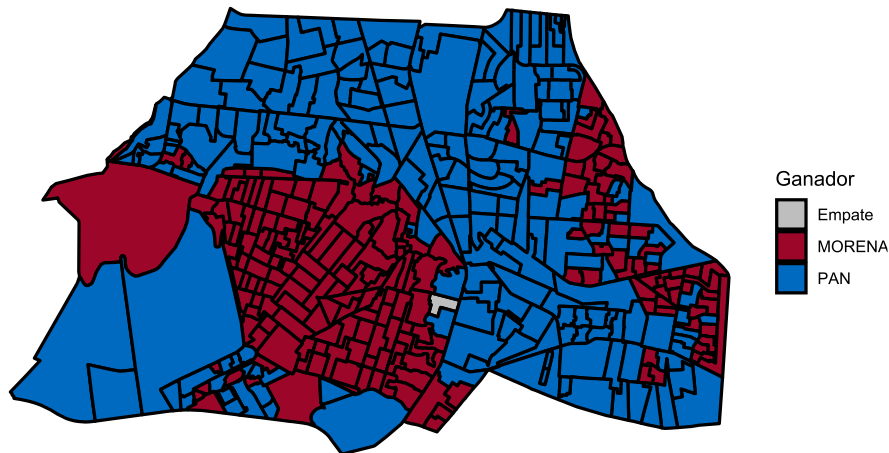


Elaborado por Alberto Reyes Briseño

Comparamos el mapa de Coyoacán anterior con el que nos representa al ganador (mayoría relativa) en cada sección electoral.

Ganador relativo por sección electoral en la
alcaldía Coyoacán

Fuente: Datos abiertos INE



Elaborado por Alberto Reyes Briseño

Concluimos que en las secciones electorales donde hay mayor porcentaje de población menor a los 29 MORENA gana, por el contrario, el PAN se debilita.

Calcularemos la probabilidad de que el 25% o más de los votos en cada sección electoral sea de menores de 29 años. Esto lo lograremos calculando con la siguiente función de probabilidad acumulada.

$$P(X \geq .25 * \text{votostotales}) = 1 - \sum_{x=0}^{.25 * \text{votostotales} - 1} \binom{\text{Votostotales}}{x} \left(\frac{\text{jovenes}}{\text{padron}} \right)^x \left(1 - \frac{\text{jovenes}}{\text{padron}} \right)^{\text{Votostotales} - x}$$

```
1 prob_voto_joven <- comparacion_votos
2 prob_voto_joven <- prob_voto_joven |>
3 mutate(Proporcion_menores29 = Menores_29/PADRON_ELECTORAL) |>
4 select(SECCION, LN.x, PAD.x, ID.y, entidad, distrito_local,
5         distrito_federal, PADRON_ELECTORAL, Menores_29, MUNICIPIO,
6 municipio,
7         TOTAL_VOTOS, LISTA_NOMINAL, geometry,
8 Proporcion_menores29)
9 #Sacamos el promedio
10 mean(prob_voto_joven$Proporcion_menores29)
11 ## función binomial
12
13
14 f2 <- function(padron, jovenes, votos){
15   porcentaje_limite <- 0.25
16   votos_totales <- votos
17   votos_jovenes_minimo <- ceiling(porcentaje_limite *
18 votos_totales)
19
```

```

20 prob_acumulada <- 1 - pbinom(votos_jovenes_minimo - 1,
21 votos_totales, jovenes / padron)
22
23 return(prob_acumulada)
24 }
25
26 f2(1555, 390, 82)
27
28 #ejemplo probabilidad_seccion(.3, 1555, 390, 821)
29 #hacemos el for para sumar
30
31 #probabilidad_acumulada<- function (padron, menores29,
32 votos_promedio){
33 #   suma <- 0
34 #   for(i in c(.2, .3 , .4, .5, .6 , .7, .8, .9, 1)){
35 #     suma <- suma + probabilidad_seccion(i, padron, menores29,
36 votos_promedio)
37 #   }
38
39 #return (suma)
40 #}

#probabilidad_acumulada(1555, 390, 82)

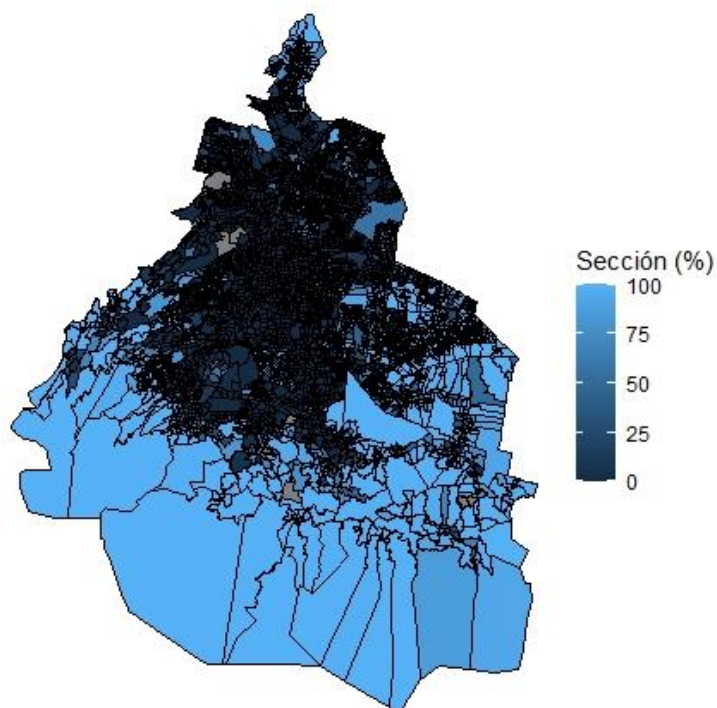
prob_voto_joven <- prob_voto_joven |>
  mutate(probabilidad_acumulada_jovenes_voto =
    (f2(PADRON_ELECTORAL, Menores_29, TOTAL_VOTOS))*100)

```

Y por último visualizaremos la información en un mapa que muestre todas las secciones electorales de la CDMX y, en particular, las secciones de Coyoacán con su respectivo porcentaje de probabilidad.

Probabilidad de que más del 25% de los votos (CDMX) sean de Menores de 29 años por sección electoral

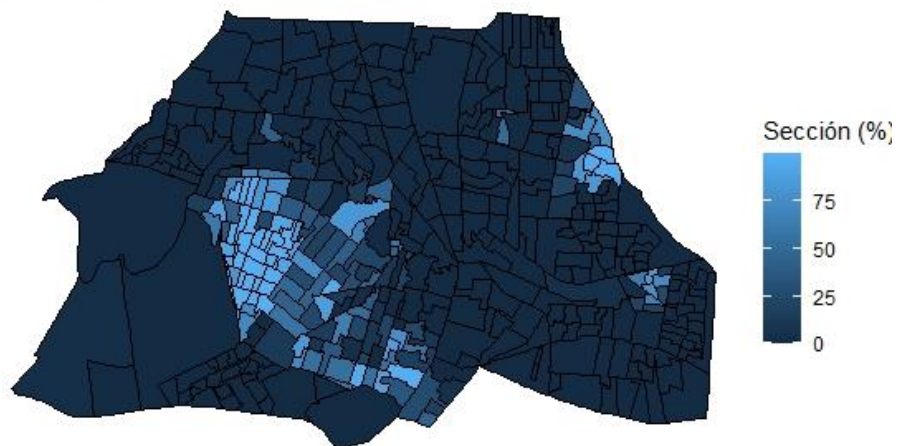
Fuente: Datos abiertos INE



Elaborado por Alberto Reyes Briseño

Probabilidad de que más del 25% de los votos (Coyoacán) sean de Menores de 29 años por sección electoral

Fuente: Datos abiertos INE



Elaborado por Alberto Reyes Briseño

Variables utilizadas

#	x
1	acu
2	asimetria
3	base
4	base_comp
5	base_ponderada
6	cat_secciones
7	cdf
8	colores
9	e2000
10	e2000cm
11	e2006
12	e2006cm
13	e2012
14	e2012cm
15	e2018
16	e2018cm
17	e2021
18	e2021cm
19	e21
20	e21cm
21	edades_sexo
22	escolaridad
23	Fa
24	frecuencia_voto
25	funh
26	funm
27	grado
28	grado_masa
29	m
30	mapa_data
31	mapa_dbf
32	mapa_entidad_9_coyo
33	mapa_entidad_9_data
34	mapa_shp
35	mapita
36	masa_con
37	masa_sexo
38	mean_dto
39	med_con

40	mh
41	mm
42	p
43	p2
44	par
45	prob_votos_seccion
46	sum_parti
47	summarise_dtos

Información de la sesión en R

R version 4.2.2 (2022-10-31 ucrt)

Platform: x86_64-w64-mingw32/x64 (64-bit)

Running under: Windows 10 x64 (build 19045)

Matrix products: default

locale:

[1] LC_COLLATE=Spanish_Mexico.utf8 LC_CTYPE=Spanish_Mexico.utf8 LC_MONETARY=Spanish_Mexico.utf8

[4] LC_NUMERIC=C LC_TIME=Spanish_Mexico.utf8

attached base packages:

[1] stats graphics grDevices utils datasets methods base

other attached packages:

[1] sf_1.0-12 foreign_0.8-83 rgdal_1.6-6 sp_1.6-0 ggExtra_0.10.0
 [6] xlsx_0.6.5 moments_0.14.1 gridExtra_2.3 readxl_1.4.2 esquisse_1.1.2
 [11] lubridate_1.9.2 forcats_1.0.0 stringr_1.5.0 dplyr_1.1.0 purrr_1.0.1
 [16] readr_2.1.4 tidyr_1.3.0 tibble_3.1.8 ggplot2_3.4.1 tidyverse_2.0.0
 [21] RColorBrewer_1.1-3

loaded via a namespace (and not attached):

[1] nlme_3.1-160 reactable_0.4.4 shinybusy_0.3.1 bit64_4.0.5 phosphoricons_0.1.2
 [6] tools_4.2.2 bslib_0.4.2 utf8_1.2.2 R6_2.5.1 KernSmooth_2.23-20
 [11] DBI_1.1.3 mgcv_1.8-41 colorspace_2.1-0 withr_2.5.0 tidyselect_1.2.0
 [16] bit_4.0.5 curl_5.0.0 compiler_4.2.2 cli_3.6.0 isoband_0.2.7
 [21] labeling_0.4.2 sass_0.4.5 scales_1.2.1 classInt_0.4-9 proxy_0.4-27
 [26] digest_0.6.31 rio_0.5.29 pkgconfig_2.0.3 htmltools_0.5.4 fastmap_1.1.1
 [31] htmlwidgets_1.6.2 rlang_1.0.6 rstudioapi_0.14 shiny_1.7.4 jquerylib_0.1.4
 [36] generics_0.1.3 farver_2.1.1 jsonlite_1.8.4 vroom_1.6.1 zip_2.2.2
 [41] magrittr_2.0.3 Matrix_1.5-1 Rcpp_1.0.10 munsell_0.5.0 fansi_1.0.4
 [46] lifecycle_1.0.3 stringi_1.7.12 MASS_7.3-58.1 grid_4.2.2 parallel_4.2.2
 [51] datamods_1.4.0 promises_1.2.0.1 crayon_1.5.2 miniUI_0.1.1.1 lattice_0.20-45
 [56] haven_2.5.2 splines_4.2.2 xlsxjars_0.6.1 hms_1.1.2 pillar_1.8.1
 [61] glue_1.6.2 data.table_1.14.8 vctrs_0.5.2 tzdb_0.3.0 httpuv_1.6.9
 [66] cellranger_1.1.0 gtable_0.3.1 cachem_1.0.7 openxlsx_4.2.5.2 mime_0.12
 [71] xtable_1.8-4 e1071_1.7-13 later_1.3.0 class_7.3-20 viridisLite_0.4.1
 [76] rJava_1.0-6 units_0.8-2 writexl_1.4.2 shinyWidgets_0.7.6 timechange_0.2.0
 [81] ellipsis_0.3.2