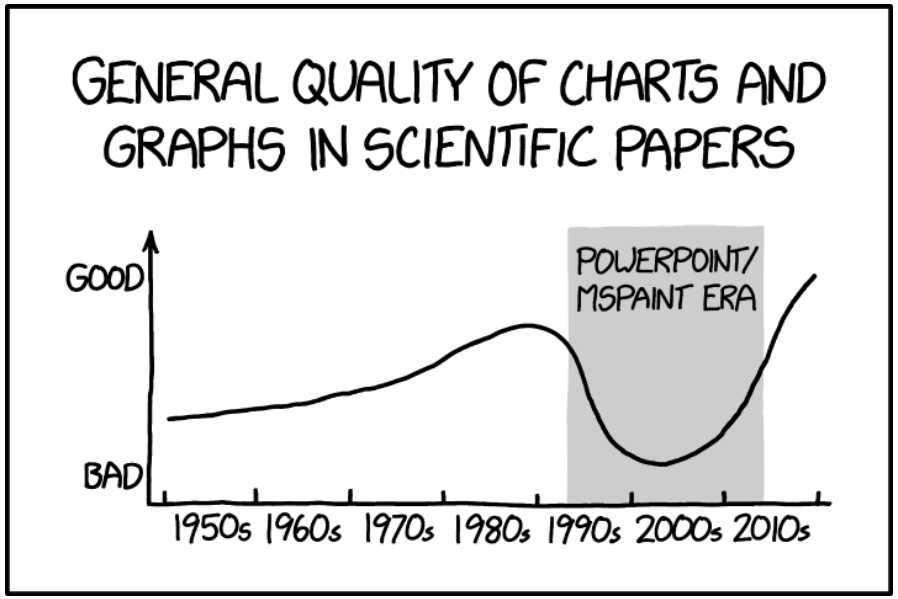


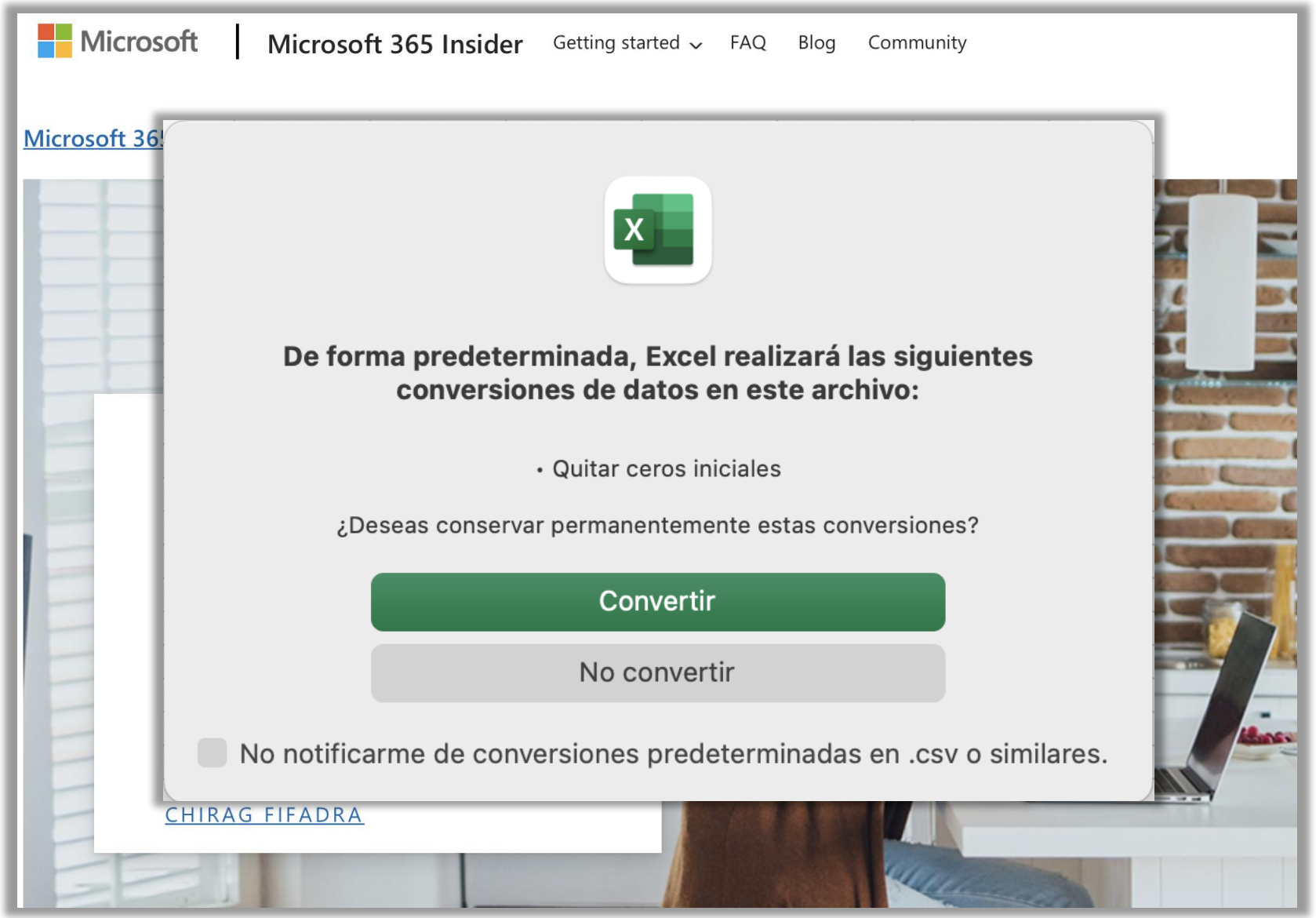
INSP/ESPM
Escuela de Salud Pública de México

Tidyverse {ggplot2}

José Luis Texcalac Sangrador

Procesamiento y visualización de datos espaciales en R

INSP/ESPM
Escuela de Salud Pública de México



id	ciudad	estado	fecha	poblacion	densidad	fertilidad	mortalidad	inmigracion
1	New York	New York	08/12/2010	8 175 133	5862	2,05	8,20	15,30
2	Chicago	Illinois	21-dec-2010	2 853 114	11 841,8	2,10	7,80	10,20
3	Houston	Texas	01/12/2010	2 099 451	3501,5	1,98	8,05	13,28
3	Houston	Texas	01/12/2010	2 099 451	3501,5	1,98	8,05	13,28
4	Los Ángeles	Cabifornia	04/12/2010	23 792 621	8092,3	2,13	8,10	15,02
A	Filadelfia	Pensilvania	03/12/2010	1 526 006		2,01	7,90	14,78

Caso duplicado

Caso no válido

Acentos

Caso erróneo

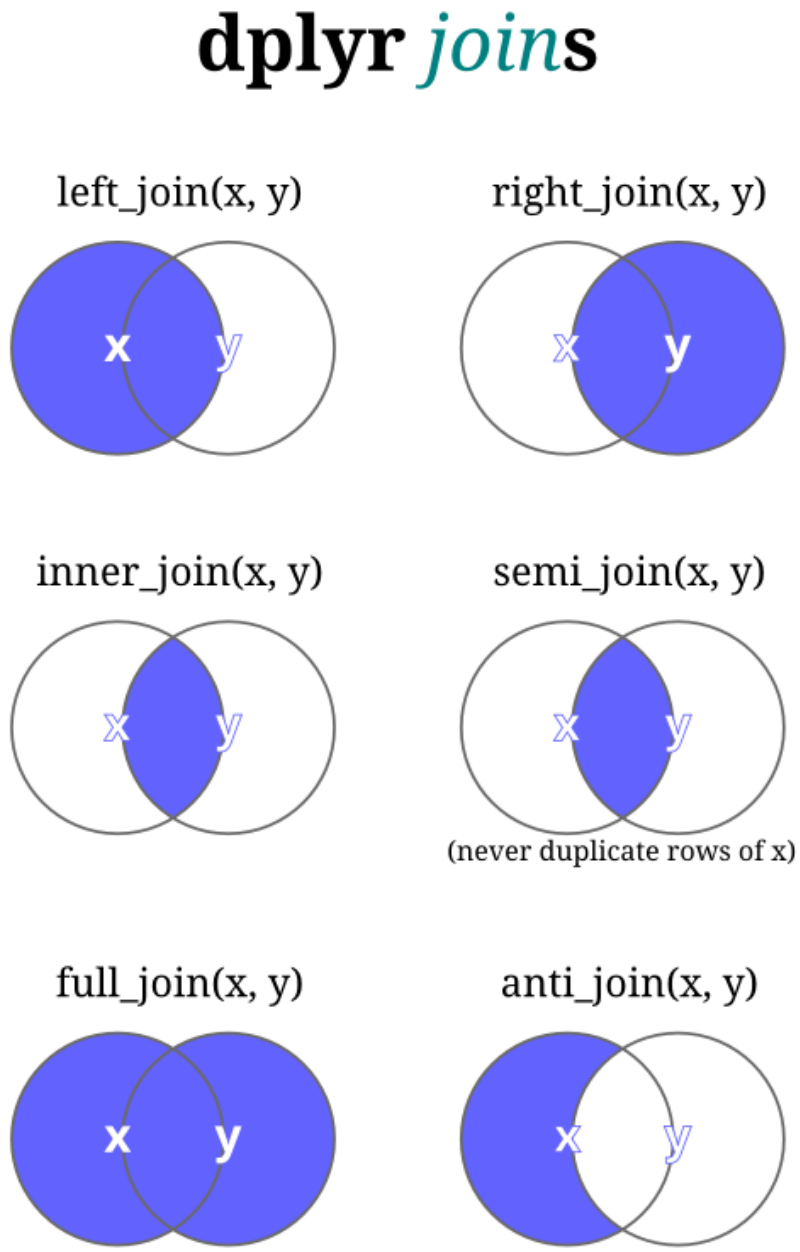
Error de formato

Caso atípico (¿error?)

Valor perdido

INSP/ESPM
Escuela de Salud Pública de México

Unión de mallas de datos



cdmx

mun_cve	mun_nom	defun	pob
09002	Azcapotzalco	5860	432205
09003	Coyoacán	7242	614447
09006	Iztapalapa	21334	1835486
09012	Tlalpan	6456	699928
09015	Cuauhtémoc	7008	545884

sun

mun_cve	cve_sun	cve_nom	pob
09002	M09.01	Valle de México	400567
09009	M09.01	Valle de México	137931
09012	M09.01	Valle de México	667260
09013	M09.01	Valle de México	415541

```
inner_join(cdmx, sun, by = "mun_cve")
```

mun_cve	mun_nom	defun	pob	cve_sun	nom_sun	pob_sun
09002	Azcapotzalco	5860	432205	M09.01	Valle de México	400567
09012	Tlalpan	6456	699928	M09.01	Valle de México	667260

```
left_join(cdmx, sun, by = "mun_cve")
```

mun_cve	mun_nom	defun	pob	cve_sun	nom_sun	pob_sun
09002	Azcapotzalco	5860	432205	M09.01	Valle de México	400567
09003	Coyoacán	7242	614447	NA	NA	NA
09006	Iztapalapa	21334	1835486	NA	NA	NA
09012	Tlalpan	6456	699928	M09.01	Valle de México	667260
09015	Cuauhtémoc	7008	545884	NA	NA	NA

cdmx

mun_cve	mun_nom	defun	pob
09002	Azcapotzalco	5860	432205
09003	Coyoacán	7242	614447
09006	Iztapalapa	21334	1835486
09012	Tlalpan	6456	699928
09015	Cuauhtémoc	7008	545884

sun

mun_cve	cve_sun	cve_nom	pob
09002	M09.01	Valle de México	400567
09009	M09.01	Valle de México	137931
09012	M09.01	Valle de México	667260
09013	M09.01	Valle de México	415541

```
right_join(cdmx, sun, by = "mun_cve")
```

mun_cve	mun_nom	defun	pob	cve_sun	nom_sun	pob_sun
09002	Azcapotzalco	5860	432205	M09.01	Valle de México	400567
09012	Tlalpan	6456	699928	M09.01	Valle de México	667260
09009	NA	NA	NA	M09.01	Valle de México	137931
09013	NA	NA	NA	M09.01	Valle de México	415541

```
anti_join(cdmx, sun, by = "mun_cve")
```

mun_cve	mun_nom	defun	pob
09003	Coyoacán	7242	614447
09006	Iztapalapa	21334	1835486
09015	Cuauhtémoc	7008	545884

cdmx

mun_cve	mun_nom	defun	pob
09002	Azcapotzalco	5860	432205
09003	Coyoacán	7242	614447
09006	Iztapalapa	21334	1835486
09012	Tlalpan	6456	699928
09015	Cuauhtémoc	7008	545884

sun

mun_cve	cve_sun	cve_nom	pob
09002	M09.01	Valle de México	400567
09009	M09.01	Valle de México	137931
09012	M09.01	Valle de México	667260
09013	M09.01	Valle de México	415541

```
full_join(cdmx, sun, by = "mun_cve")
```

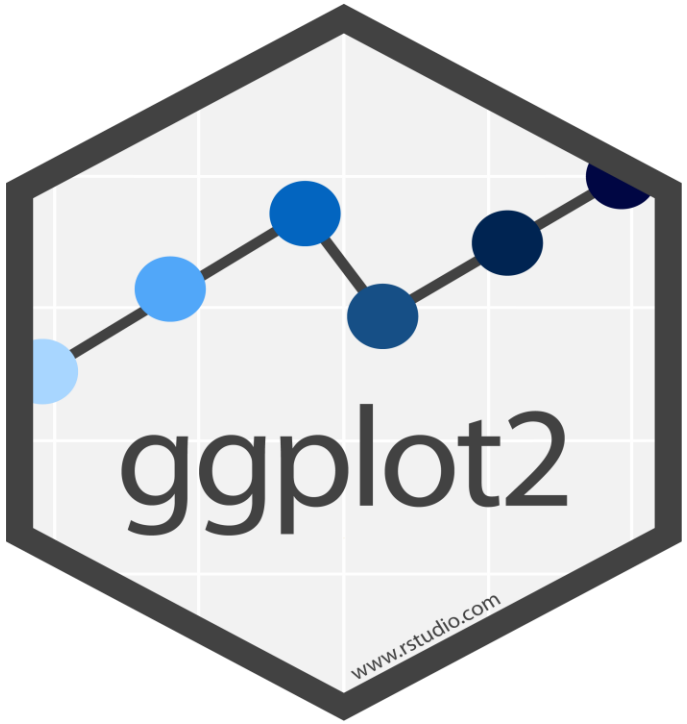
mun_cve	mun_nom	defun	pob	cve_sun	nom_sun	pob_sun
09002	Azcapotzalco	5860	432205	M09.01	Valle de México	400567
09003	Coyoacán	7242	614447	NA	NA	NA
09006	Iztapalapa	21334	1835486	NA	NA	NA
09012	Tlalpan	6456	699928	M09.01	Valle de México	667260
09015	Cuauhtémoc	7008	545884	NA	NA	NA
09009	NA	NA	NA	M09.01	Valle de México	137931
09013	NA	NA	NA	M09.01	Valle de México	415541

"The greatest value of a picture is when it forces us to notice what we never expected to see."

– John Tukey

INSP/ESPM
Escuela de Salud Pública de México

Tidyverse {ggplot}



- Misma función que el pipe
- Sustituye al pipe
- Siempre al final de la línea, NO al inicio

Malla de datos

```
ggplot(data = dataset) +
  geom_col(mapping = aes(x = x_var,    y = y_var))
```

Tipo de geometría

aesthetics

Columna del eje X

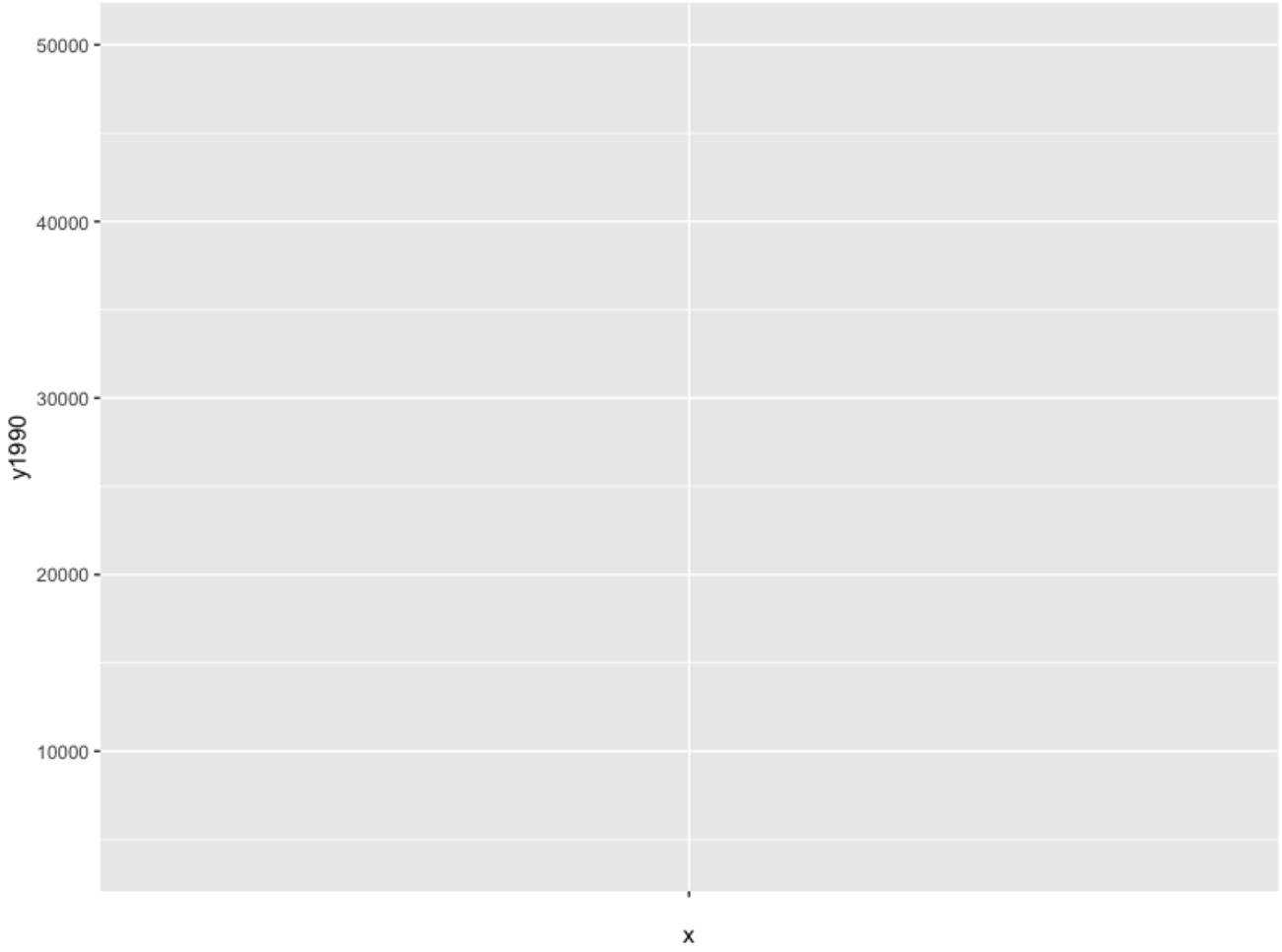
Columna del eje Y

Forma resumida de redacción

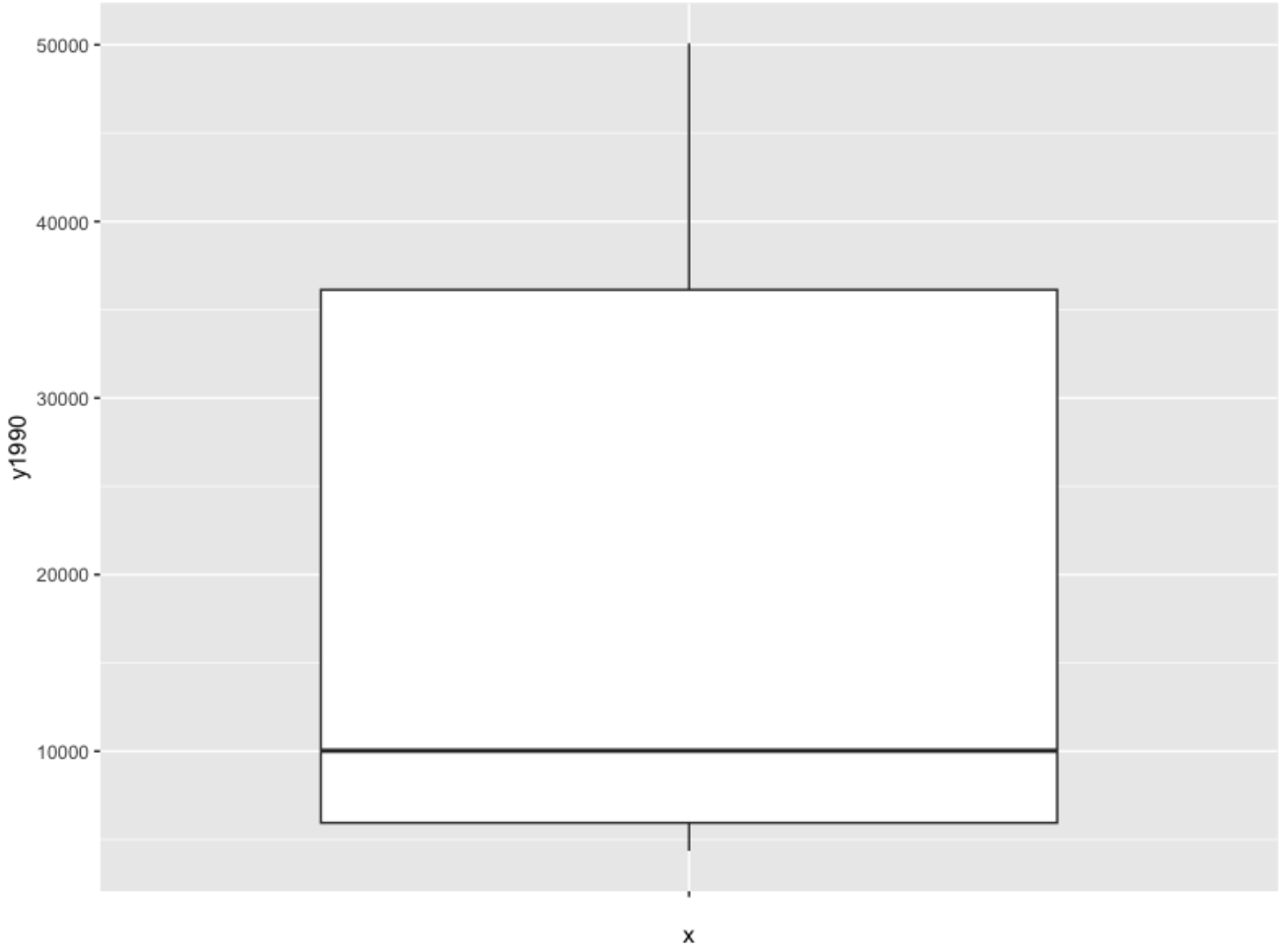
```
ggplot(dataset) +
  geom_col(aes(x_var, y_var))
```

ggplot(data = defun_w)

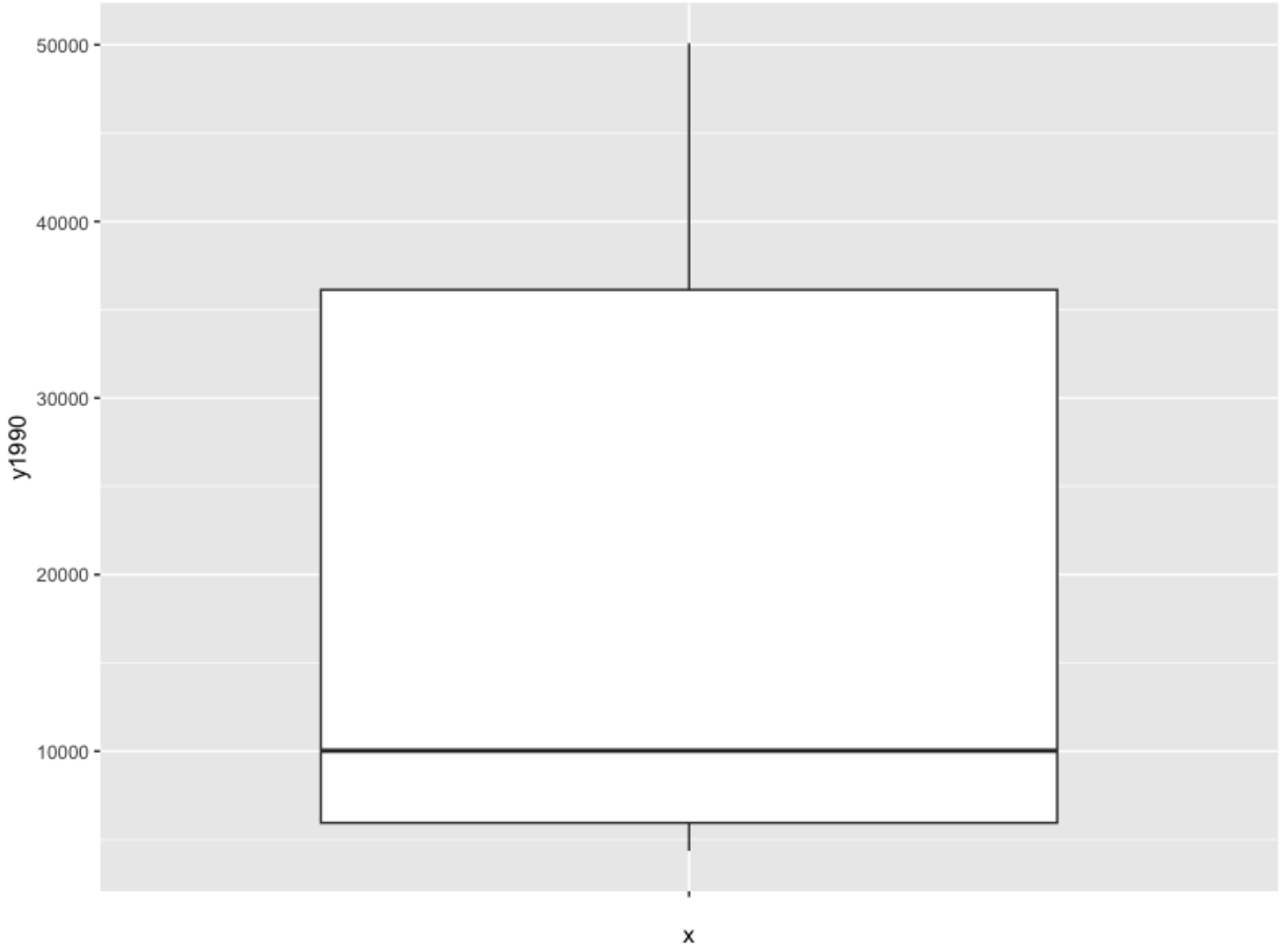
```
ggplot(data = defun_w, aes(x = "", y1990))
```



```
ggplot(data = defun_w, aes(x = "", y1990)) +
    geom_boxplot( )
```



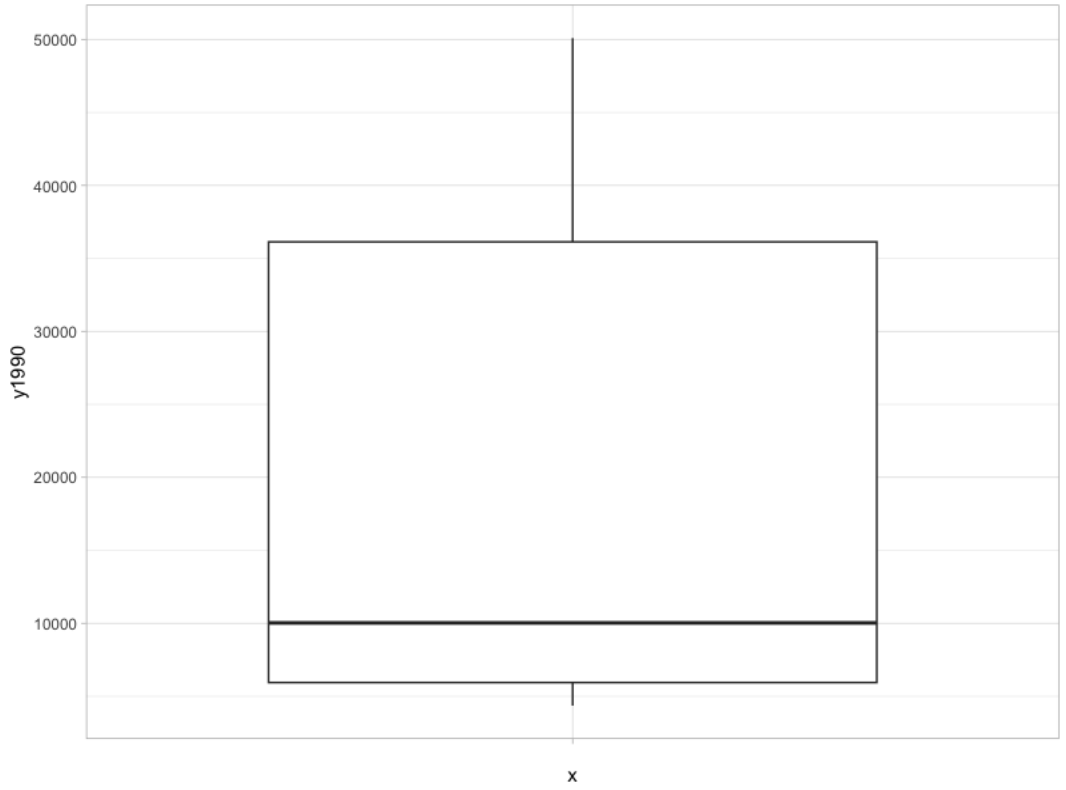
```
ggplot(data = defun_w) +
    geom_boxplot(aes(x = "", y1990))
```



Temas en gráficos

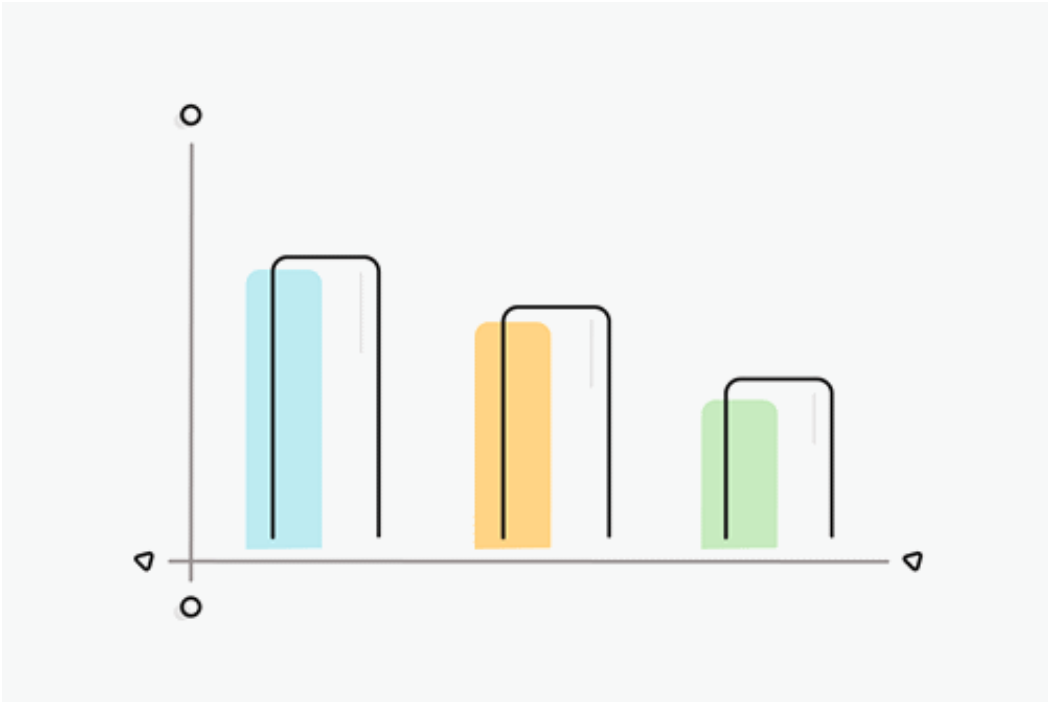
- theme_grey()
- theme_gray()
- theme_bw()
- theme_linedraw()
- theme_dark()
- theme_light()
- theme_minimal()

```
ggplot(data = defun_w) +
  geom_boxplot(aes("", y1990)) +
  theme_light()
```



Su turno. . .

- Replique el gráfico previo usando los distintos temas disponibles.



https://ggplot2tor.com/theme/

Colores en R

- Colores por nombre y por hex code:

https://rpubs.com/kylewbrown/r-colors

INSP/ESPM
Escuela de Salud Pública de México

Colores por nombre y hexcode

white	coral4	deepskyblue	gray28	gray88	grey40	grey100	lightpink2	mistyrose2	plum	slategray2
aliceblue	cornflowerblue	deepskyblue1	gray29	gray89	grey41	honeydew	lightpink3	mistyrose3	plum1	slategray3
antiquewhite	cornsilk	deepskyblue2	gray30	gray90	grey42	honeydew1	lightpink4	mistyrose4	plum2	slategray4
antiquewhite1	cornsilk1	deepskyblue3	gray31	gray91	grey43	honeydew2	lightsalmon	moccasin	plum3	slategrey
antiquewhite2	cornsilk2	deepskyblue4	gray32	gray92	grey44	honeydew3	lightsalmon1	navajowhite	plum4	snow
antiquewhite3	cornsilk3	dimgray	gray33	gray93	grey45	honeydew4	lightsalmon2	navajowhite1	powderblue	snow1
antiquewhite4	cornsilk4	dimgrey	gray34	gray94	grey46	hotpink	lightsalmon3	navajowhite2	purple	snow2
aquamarine	cyan	dodgerblue	gray35	gray95	grey47	hotpink1	lightsalmon4	navajowhite3	purple1	snow3
aquamarine1	cyan1	dodgerblue1	gray36	gray96	grey48	hotpink2	lightseagreen	navajowhite4	purple2	snow4
aquamarine2	cyan2	dodgerblue2	gray37	gray97	grey49	hotpink3	lightskyblue	navy	purple3	springgreen
aquamarine3	cyan3	dodgerblue3	gray38	gray98	grey50	hotpink4	lightskyblue1	navyblue	purple4	springgreen1
aquamarine4	cyan4	dodgerblue4	gray39	gray99	grey51	indianred	lightskyblue2	oldlace	red	springgreen2
azure	darkblue	firebrick	gray40	gray100	grey52	indianred1	lightskyblue3	olivedrab	red1	springgreen3
azure1	darkcyan	firebrick1	gray41	green	grey53	indianred2	lightskyblue4	olivedrab1	red2	springgreen4
azure2	darkgoldenrod	firebrick2	gray42	green1	grey54	indianred3	lightslateblue			
azure3	darkgoldenrod1	firebrick3	gray43	green2	grey55	indianred4	lightslategray	olivedrab3	red4	steelblue1
azure4	darkgoldenrod2	firebrick4	gray44	green3	grey56	ivory	lightslategrey	olivedrab4	rosybrown	steelblue2
beige	darkgoldenrod3	floralwhite	gray45	green4	grey57	ivory1	lightsteelblue	orange	rosybrown1	steelblue3
bisque	darkgoldenrod4	forestgreen	gray46	greenyellow	grey58	ivory2	lightsteelblue1	orange1	rosybrown2	steelblue4
bisque1	darkgray	gainsboro	gray47	grey	grey59	ivory3	lightsteelblue2	orange2	rosybrown3	tan
bisque2	darkgreen	ghostwhite	gray48	grey0	grey60	ivory4	lightsteelblue3	orange3	rosybrown4	tan1
bisque3	darkgrey	gold	gray49	grey1	grey61	khaki	lightsteelblue4	orange4	royalblue	tan2
bisque4	darkkhaki	gold1	gray50	grey2	grey62	khaki1	lightyellow	orangered	royalblue1	tan3
black	darkmagenta	gold2	gray51	grey3	grey63	khaki2	lightyellow1	orangered1	royalblue2	tan4
blanchedalmond	darkolivegreen	gold3	gray52	grey4	grey64	khaki3	lightyellow2	orangered2	royalblue3	thistle
blue	darkolivegreen1	gold4	gray53	grey5	grey65	khaki4	lightyellow3	orangered3	royalblue4	thistle1
blue1	darkolivegreen2	goldenrod	gray54	grey6	grey66	lavender	lightyellow4	orangered4	saddlebrown	thistle2
blue2	darkolivegreen3	goldenrod1	gray55	grey7	grey67	lavenderblush	limegreen	orchid	salmon	thistle3
blue3	darkolivegreen4	goldenrod2	gray56	grey8	grey68	lavenderblush1	linen	orchid1	salmon1	thistle4
blue4	darkorange	goldenrod3	gray57	grey9	grey69	lavenderblush2	magenta	orchid2	salmon2	tomato
blueviolet	darkorange1	goldenrod4	gray58	grey10	grey70	lavenderblush3	magenta1	orchid3	salmon3	tomato1
brown	darkorange2	gray	gray59	grey11	grey71	lavenderblush4	magenta2	orchid4	salmon4	tomato2
brown1	darkorange3	gray0	gray60	grey12	grey72	lawngreen	magenta3	palegoldenrod	sandybrown	tomato3
brown2	darkorange4	gray1	gray61	grey13	grey73	lemonchiffon	magenta4	palegreen	seagreen	tomato4
brown3	darkorchid	gray2	gray62	grey14	grey74	lemonchiffon1	maroon	palegreen1	seagreen1	turquoise
brown4	darkorchid1	gray3	gray63	grey15	grey75	lemonchiffon2	maroon1	palegreen2	seagreen2	turquoise1
burlywood	darkorchid2	gray4	gray64	grey16	grey76	lemonchiffon3	maroon2	palegreen3	seagreen3	turquoise2
burlywood1	darkorchid3	gray5	gray65	grey17	grey77	lemonchiffon4	maroon3	palegreen4	seagreen4	turquoise3
burlywood2	darkorchid4	gray6	gray66	grey18	grey78	lightblue	maroon4	paleturquoise	seashell	turquoise4
burlywood3	darkred	gray7	gray67	grey19	grey79	lightblue1	mediumaquamarine	paleturquoise1	seashell1	violet
burlywood4	darksalmon	gray8	gray68	grey20	grey80	lightblue2	mediumblue	paleturquoise2	seashell2	violetred
cadetblue	darkseagreen	gray9	gray69	grey21	grey81	lightblue3	mediumorchid	paleturquoise3	seashell3	violetred1
cadetblue1	darkseagreen1	gray10	gray70	grey22	grey82	lightblue4	mediumorchid1	paleturquoise4	seashell4	violetred2
cadetblue2	darkseagreen2	gray11	gray71	grey23	grey83	lightcoral	mediumorchid2	palevioletred	sienna	violetred3
cadetblue3	darkseagreen3	gray12	gray72	grey24	grey84	lightcyan	mediumorchid3	palevioletred1	sienna1	violetred4
cadetblue4	darkseagreen4	gray13	gray73	grey25	grey85	lightcyan1	mediumorchid4	palevioletred2	sienna2	wheat
chartreuse	darkslateblue	gray14	gray74	grey26	grey86	lightcyan2	mediumpurple	palevioletred3	sienna3	wheat1
chartreuse1	darkslategray	gray15	gray75	grey27	grey87	lightcyan3	mediumpurple1	palevioletred4	sienna4	wheat2
chartreuse2	darkslategray1	gray16	gray76	grey28	grey88	lightcyan4	mediumpurple2	papayawhip	skyblue	wheat3
chartreuse3	darkslategray2	gray17	gray77	grey29	grey89	lightgoldenrod	mediumpurple3	peachpuff	skyblue1	wheat4
chartreuse4	darkslategray3	gray18	gray78	grey30	grey90	lightgoldenrod1	mediumpurple4	peachpuff1	skyblue2	whitesmoke
chocolate	darkslategray4	gray19	gray79	grey31	grey91	lightgoldenrod2	mediumseagreen	peachpuff2	skyblue3	yellow
chocolate1	darkslategrey	gray20	gray80	grey32	grey92	lightgoldenrod3	mediumslateblue	peachpuff3	skyblue4	yellow1
chocolate2	darkturquoise	gray21	gray81	grey33	grey93	lightgoldenrod4	mediumspringgreen	peachpuff4	slateblue	yellow2
chocolate3	darkviolet	gray22	gray82	grey34	grey94	lightgoldenrodyellow	mediumturquoise	peru	slateblue1	yellow3
chocolate4	deeppink	gray23	gray83	grey35	grey95	lightgray	mediumvioletred	pink	slateblue2	yellow4
coral	deeppink1	gray24	gray84	grey36	grey96	lightgreen	midnightblue	pink1	slateblue3	yellowgreen
coral1	deeppink2	gray25	gray85	grey37	grey97	lightgrey	mintcream	pink2	slateblue4	
coral2	deeppink3	gray26	gray86	grey38	grey98	lightpink	mistyrose	pink3	slategray	
coral3	deeppink4	gray27	gray87	grey39	grey99	lightpink1	mistyrose1	pink4	slategray1	

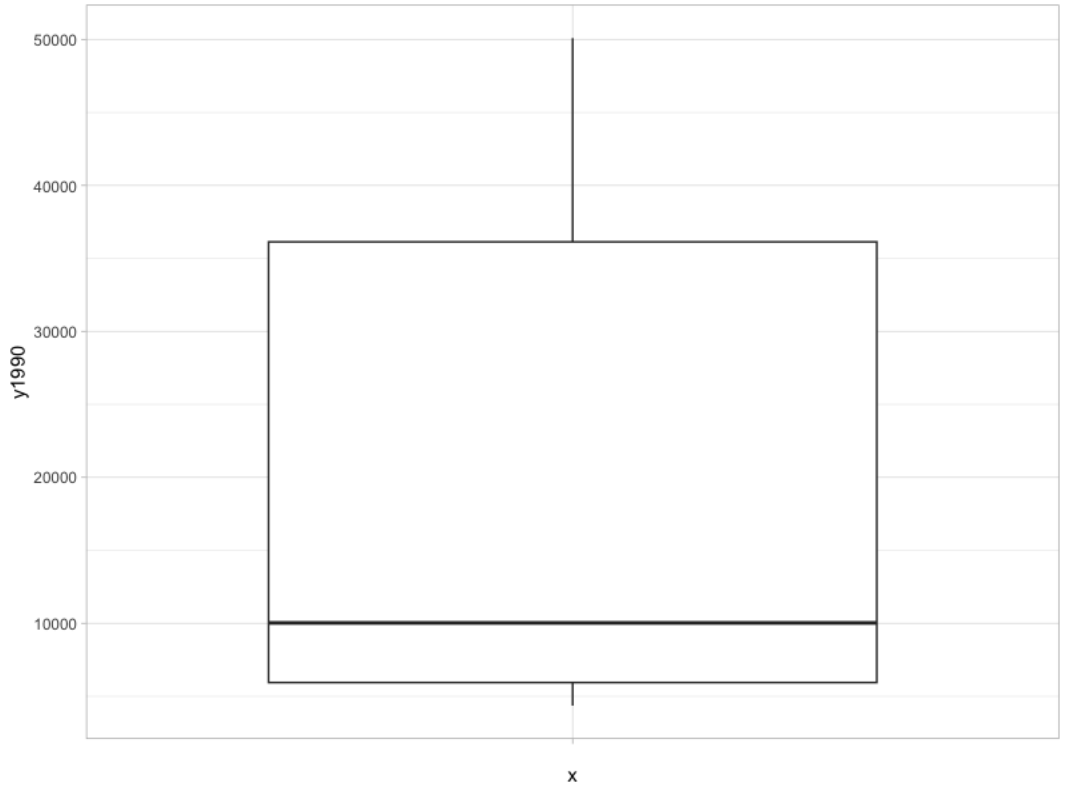
mistyrose2	plum	slategray2
mistyrose3	plum1	slategray3
mistyrose4	plum2	slategray4
moccasin	plum3	slategrey
navajowhite	plum4	snow
navajowhite1	powderblue	snow1
navajowhite2	purple	snow2
navajowhite3	purple1	snow3
navajowhite4	purple2	snow4
navy	purple3	springgreen
navyblue	purple4	springgreen1
oldlace	red	springgreen2
olivedrab	red1	springgreen3
olivedrab1	red2	springgreen4

#EED5D2	#DDA0DD	#B9D3EE
#CDB7B5	#FFBBFF	#9FB6CD
#8B7D7B	#EEAEEE	#6C7B8B
#FFE4B5	#CD96CD	#708090
#FFDEAD	#8B668B	#FFFAFA
#FFDEAD	#B0E0E6	#FFFAFA
#EECFA1	#A020F0	#EEE9E9
#CDB38B	#9B30FF	#CDC9C9
#8B795E	#912CEE	#8B8989
#000080	#7D26CD	#00FF7F
#000080	#551A8B	#00FF7F
#FDF5E6	#FF0000	#00EE76
#6B8E23	#FF0000	#00CD66
#C0FF3E	#EE0000	#008B45

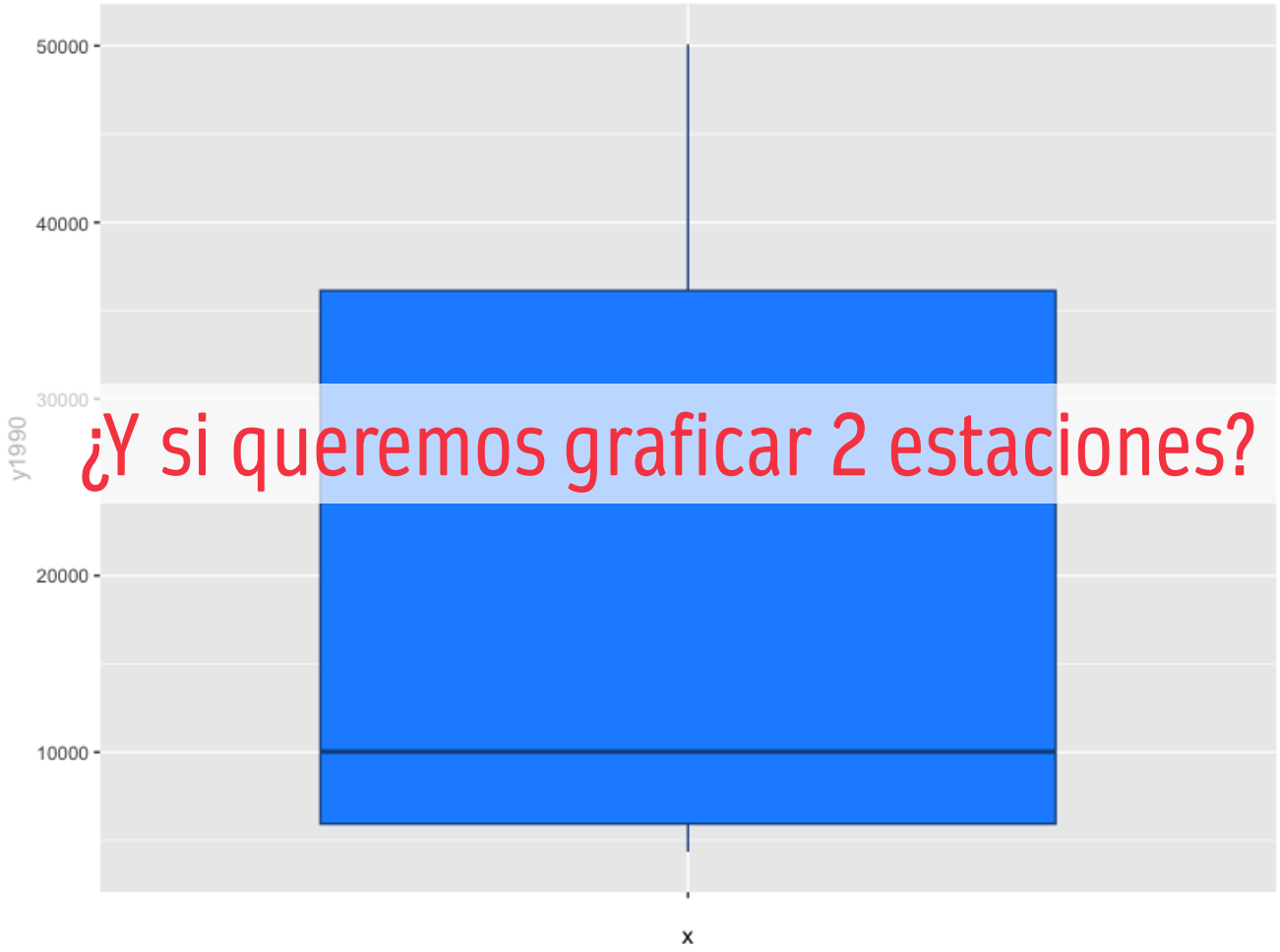
Temas en gráficos

- theme_grey()
- theme_gray()
- theme_bw()
- theme_linedraw()
- theme_dark()
- theme_light()
- theme_minimal()

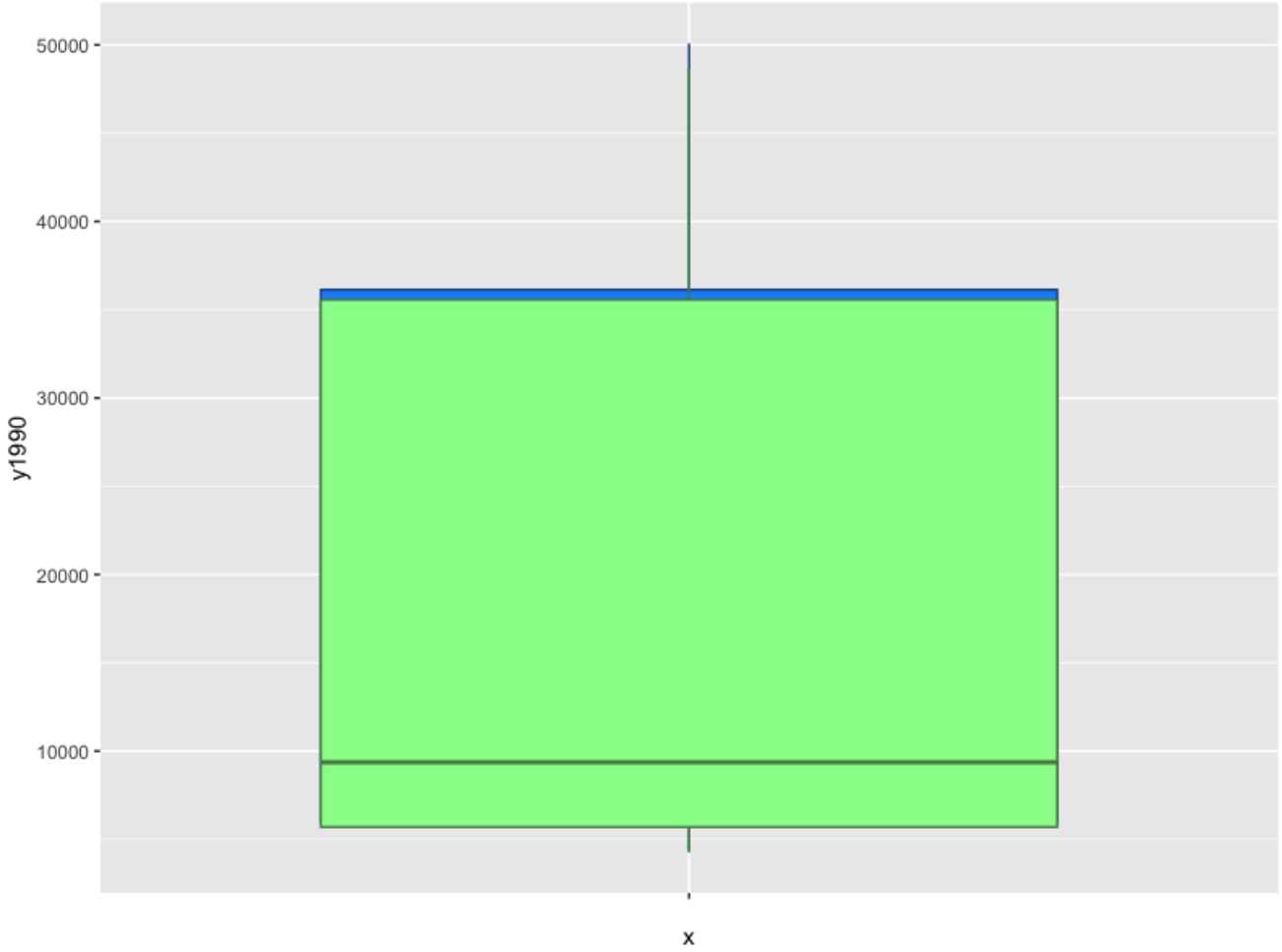
```
ggplot(data = defun_w) +
  geom_boxplot(aes("", y1990)) +
  theme_light()
```



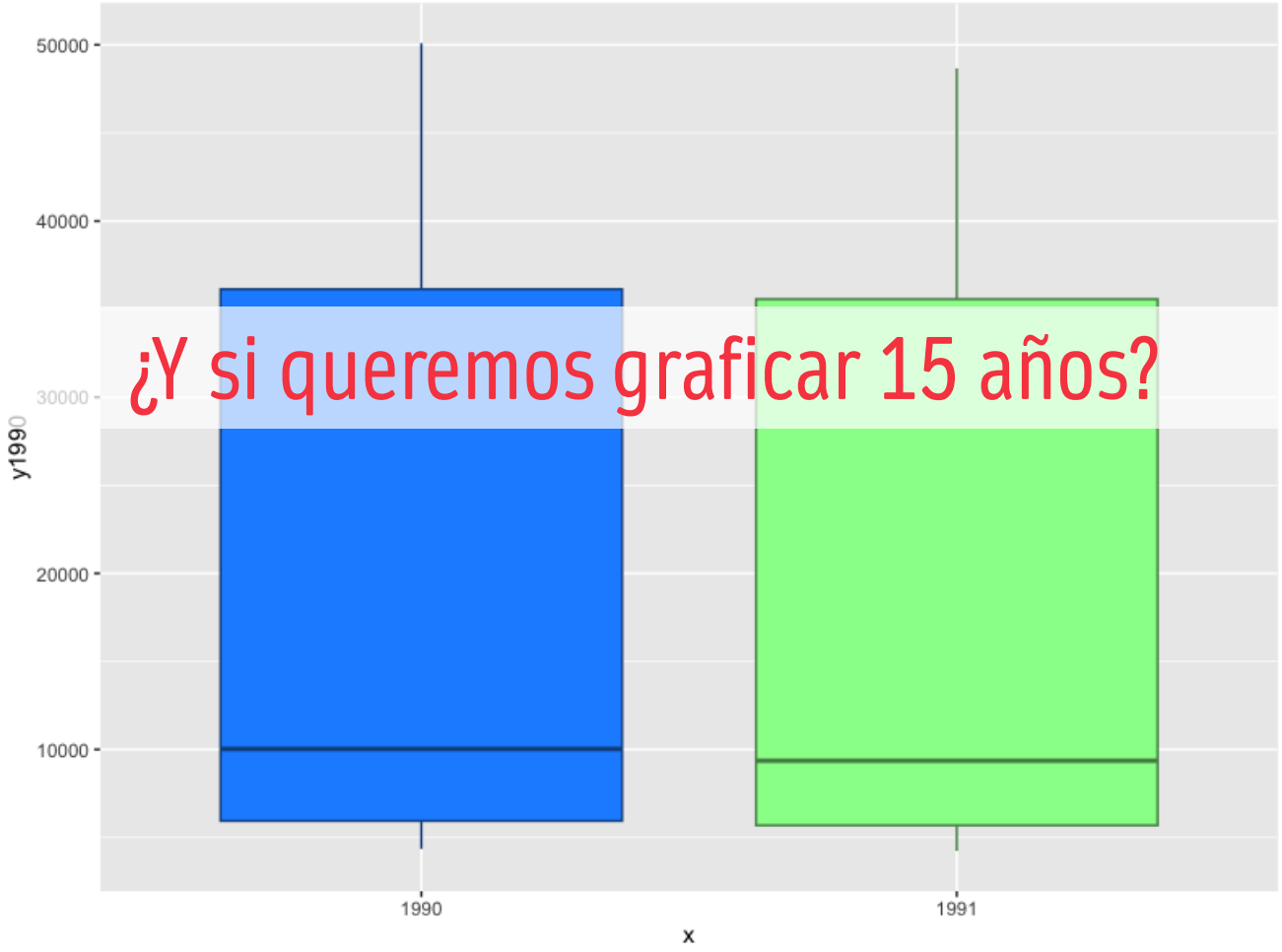
```
ggplot(data = defun_w) +
    geom_boxplot(aes(x = "", y1990), colour = "dodgerblue4", fill = "dodgerblue1")
```



```
ggplot(data = defun_w) +
    geom_boxplot(aes(x = "", y1990), colour = "dodgerblue4", fill = "dodgerblue1") +
    geom_boxplot(aes(x = "", y1991), colour = "#548B54", fill = "#98FB98")
```



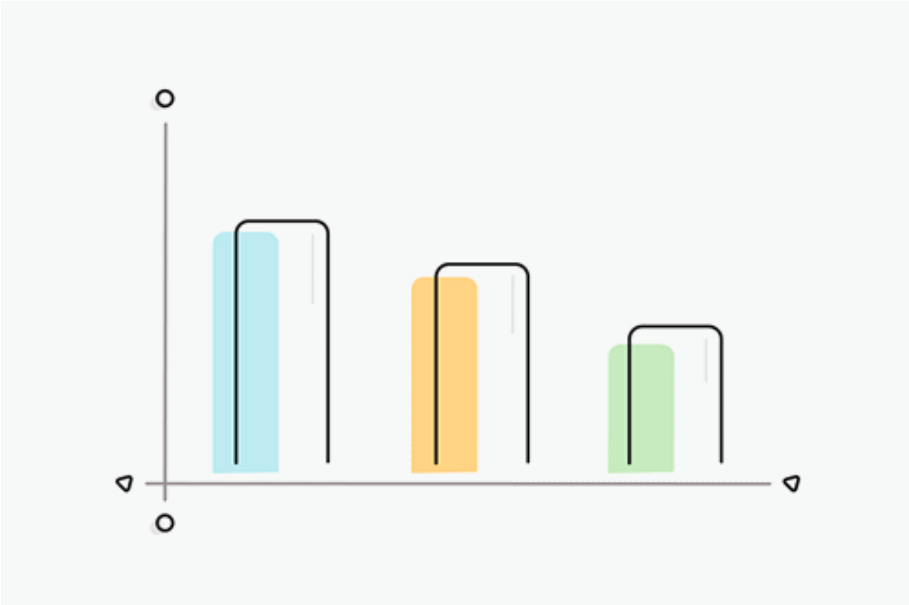
```
ggplot(data = defun_w) +
    geom_boxplot(aes(x = "1990", y1990), colour = "dodgerblue4", fill = "dodgerblue1") +
    geom_boxplot(aes(x = "1991", y1991), colour = "#548B54", fill = "#98FB98")
```



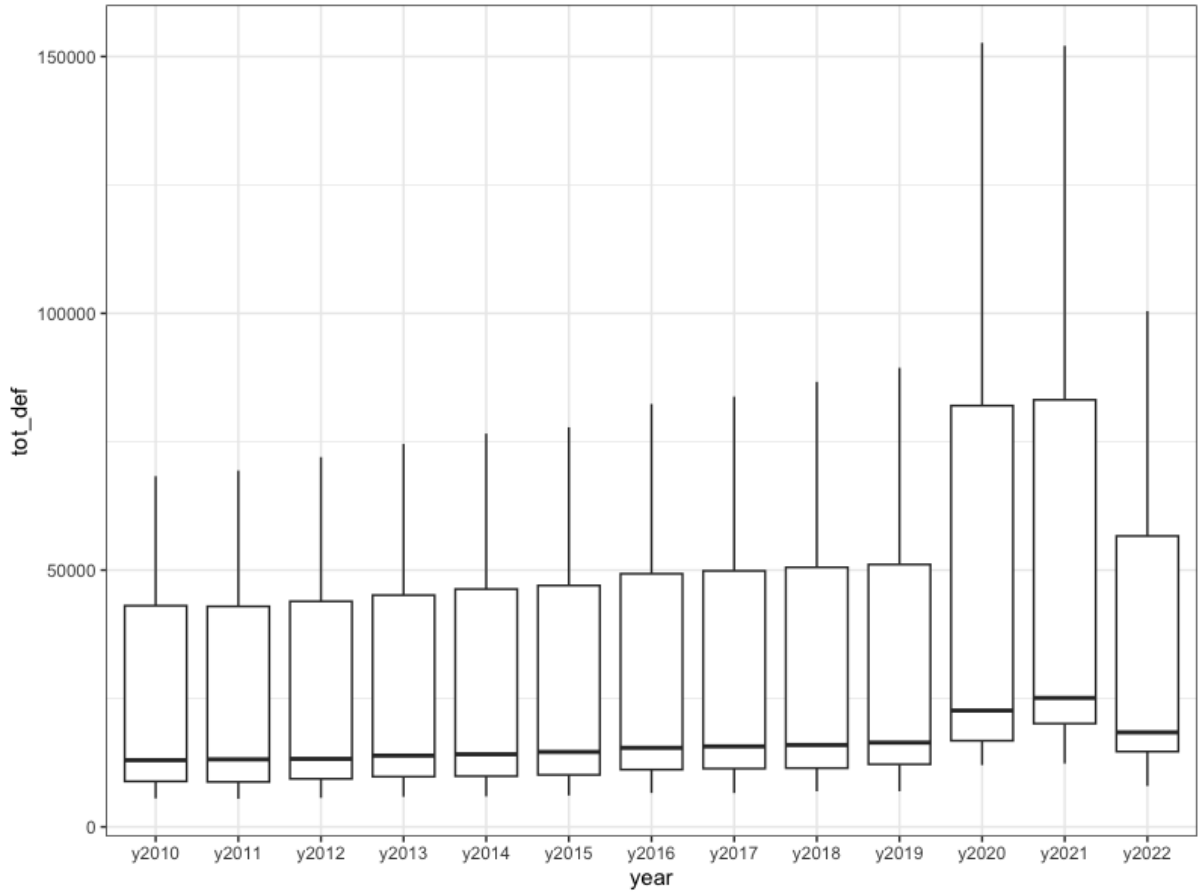
¿Y si queremos graficar 15 años?

Estructura de datos en {ggplot2}

- La graficación es más eficiente con datos estructurados en formato "long", es decir, una columna para cada variable y una fila para cada observación.
- La correcta estructura de los datos le ahorrará mucho tiempo al generar gráficos con ggplot2.
- Los gráficos de ggplot se construyen agregando capa por capa (geometrías u otros elementos).
- La graficación por capas otorga una flexibilidad y personalización de los gráficos.



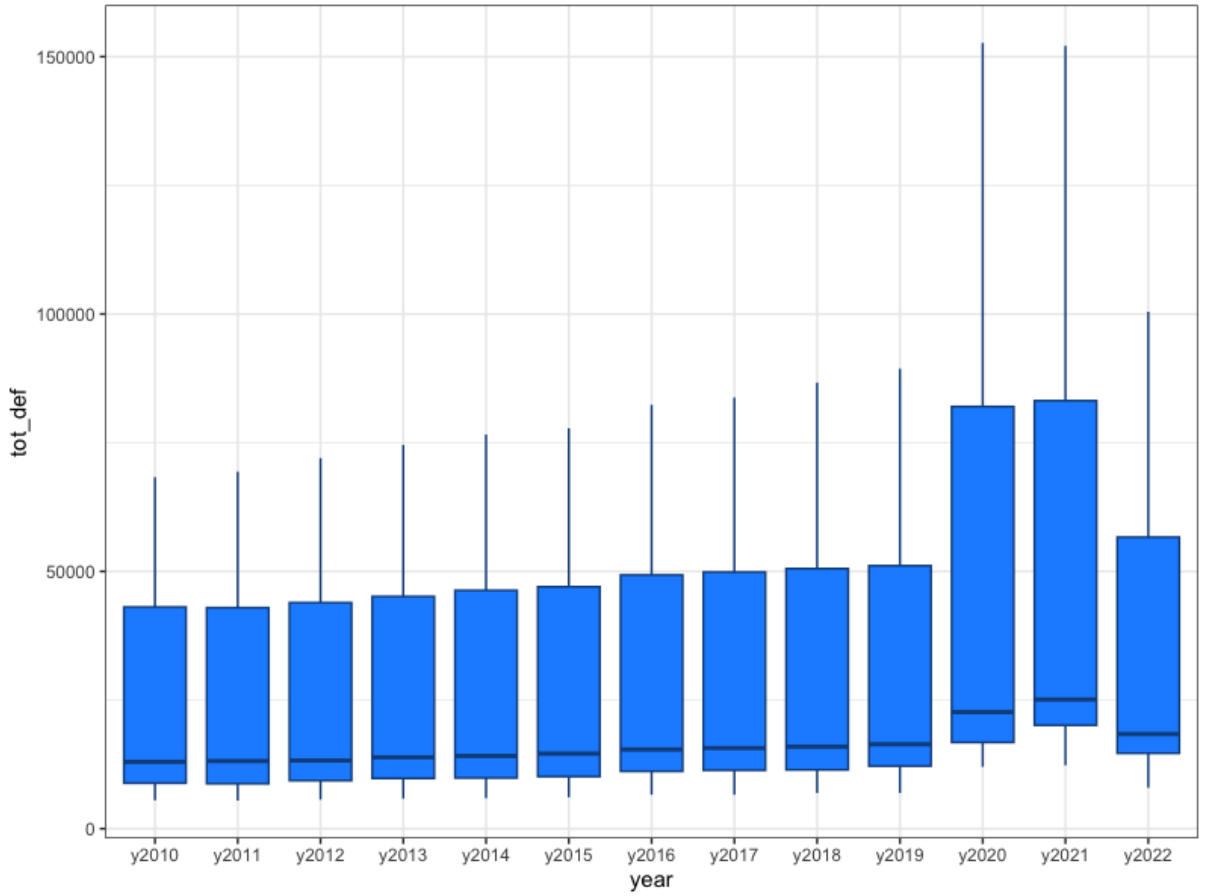
```
ggplot(data = defun_l) +
    geom_boxplot(aes(year, tot_def)) +
    theme_bw()
```



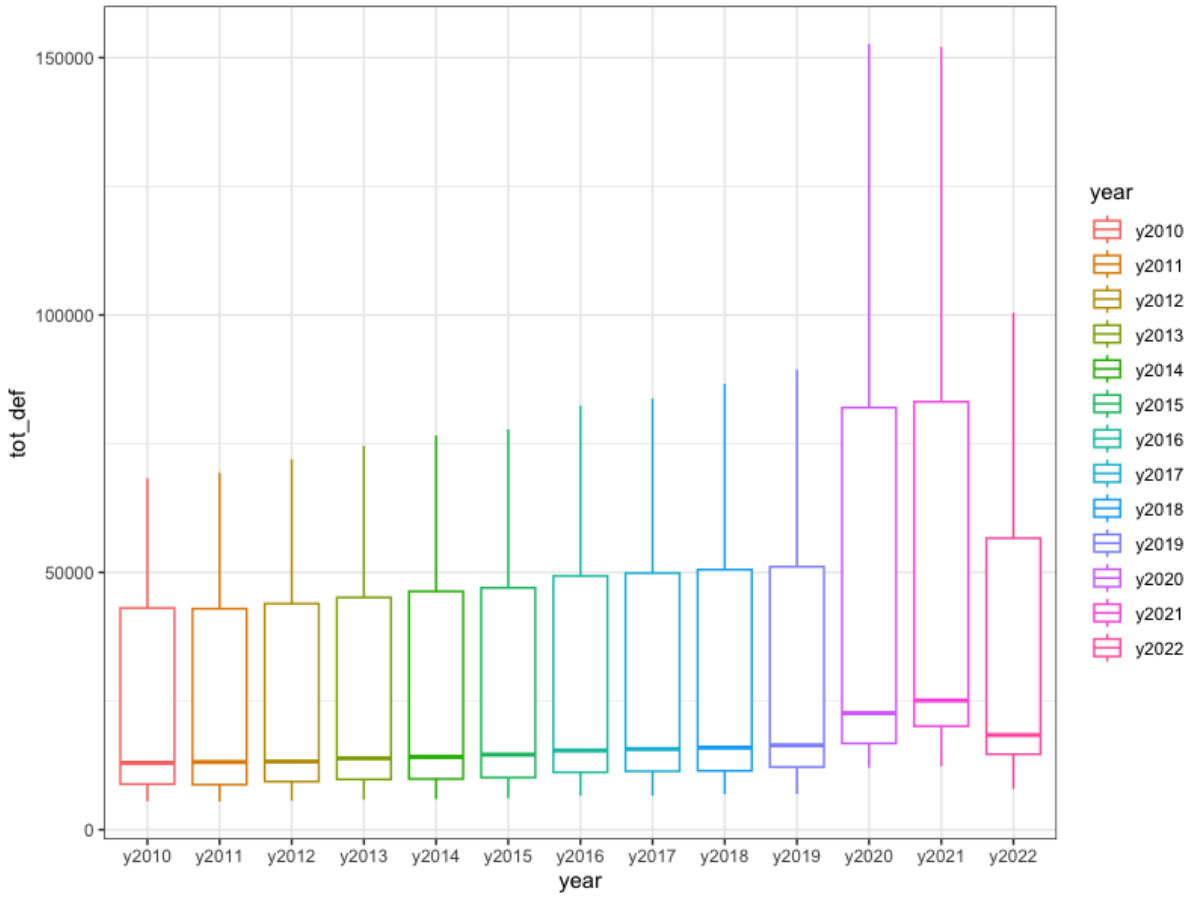
```
ggplot(data = defun_l) +
    geom_boxplot(aes(year, tot_def), colour = "dodgerblue4", fill = "dodgerblue1") +
    theme_bw()
```

Argumentos "color" y "fill"

- color: define el color de un borde.
- fill: define el color de relleno.
- Dentro de "aes" aplican un color distinto a cada elemento a graficar.
- Fuera de "aes" aplican a todos los elementos a graficar.

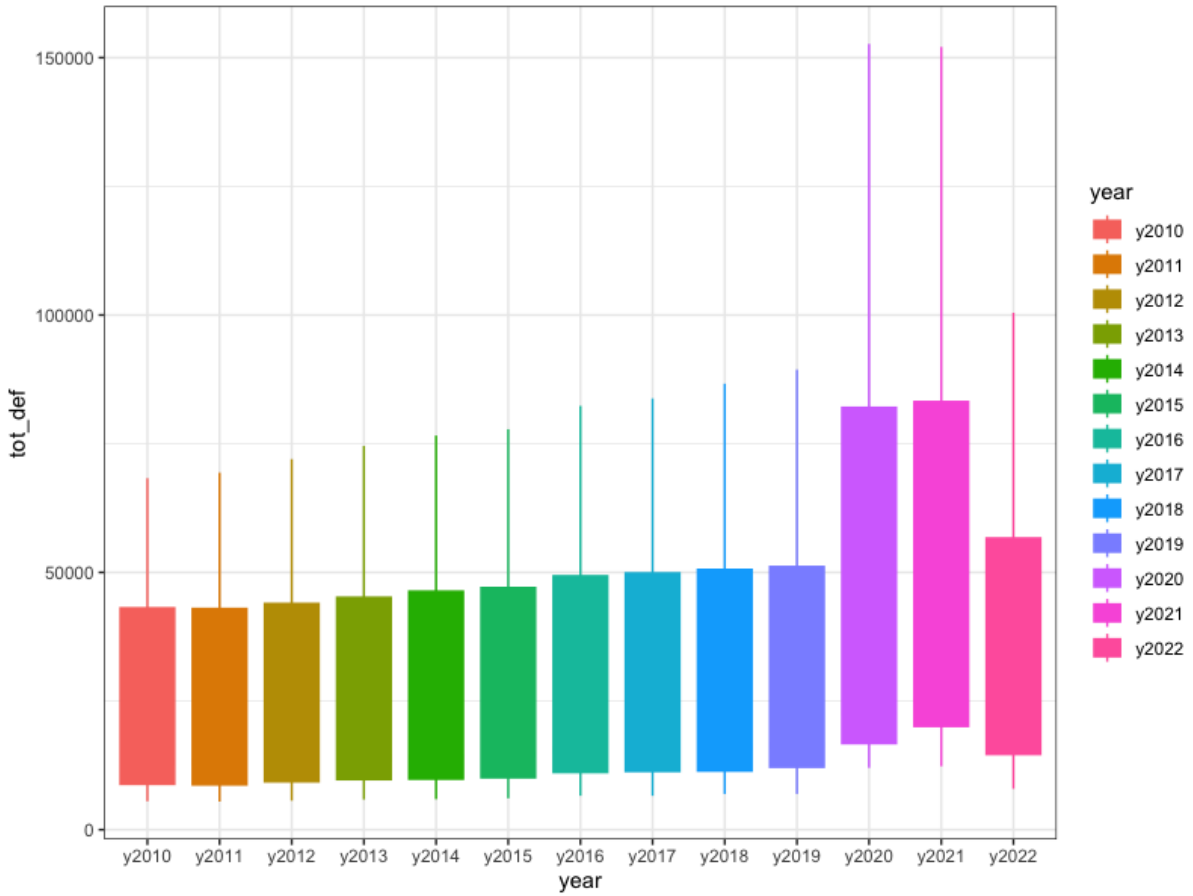


```
ggplot(data = defun_l) +
    geom_boxplot(aes(year, tot_def, color = year)) +
    theme_bw()
```



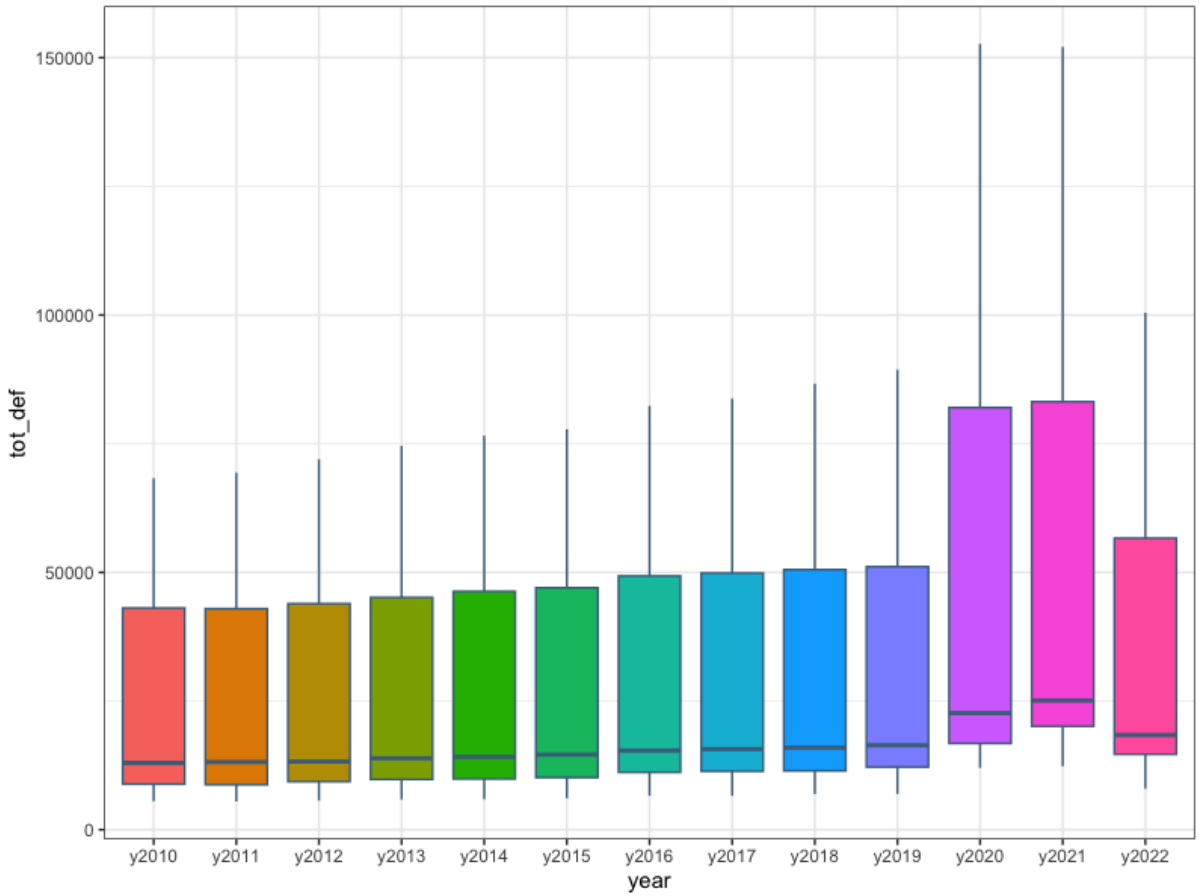
INSP/ESPM
Escuela de Salud Pública de México

```
ggplot(data = defun_l) +
    geom_boxplot(aes(year, tot_def, color = year, fill = year)) +
    theme_bw()
```



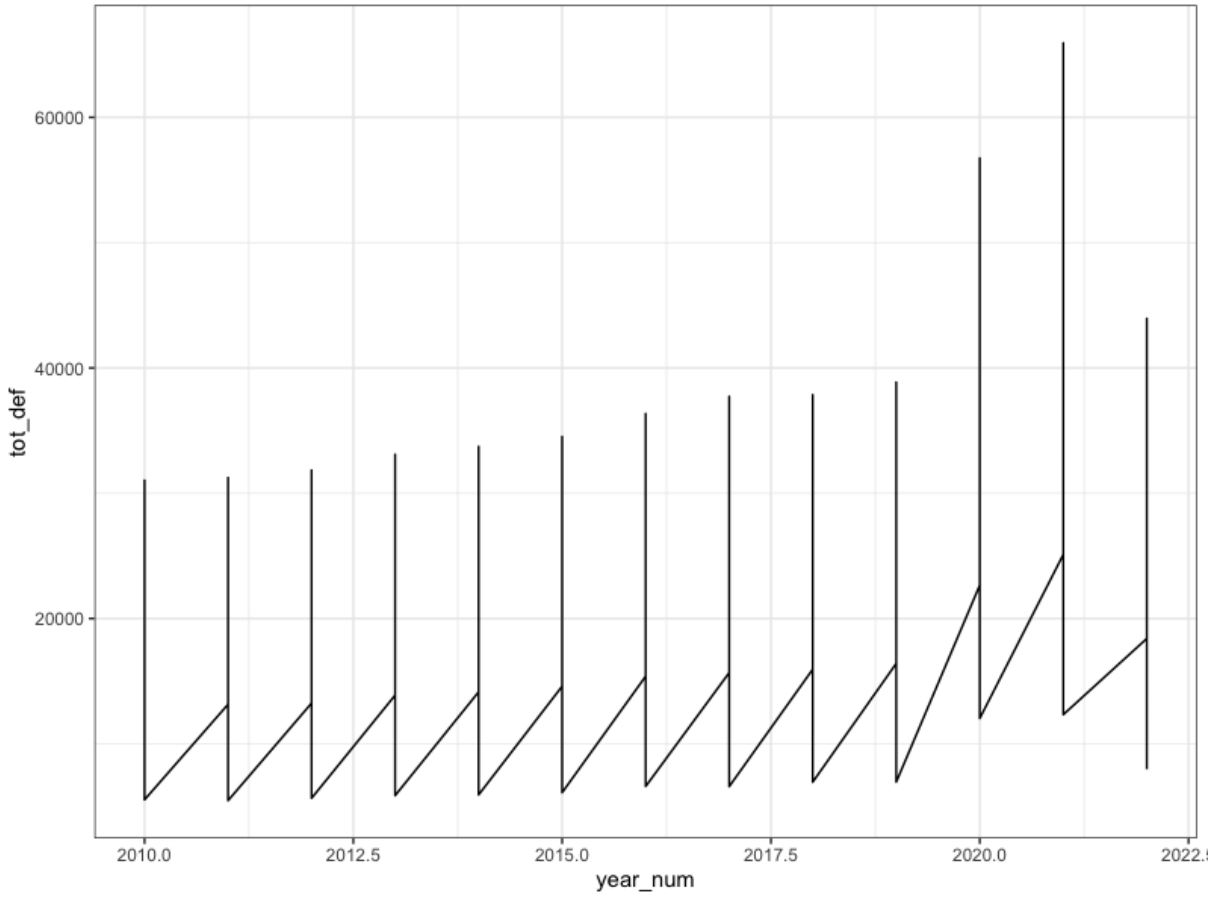
INSP/ESPM
Escuela de Salud Pública de México

```
ggplot(data = defun_l) +
    geom_boxplot(aes(year, tot_def, fill = year), color = "skyblue4") +
    theme_bw() +
    theme(legend.position = "none")
```

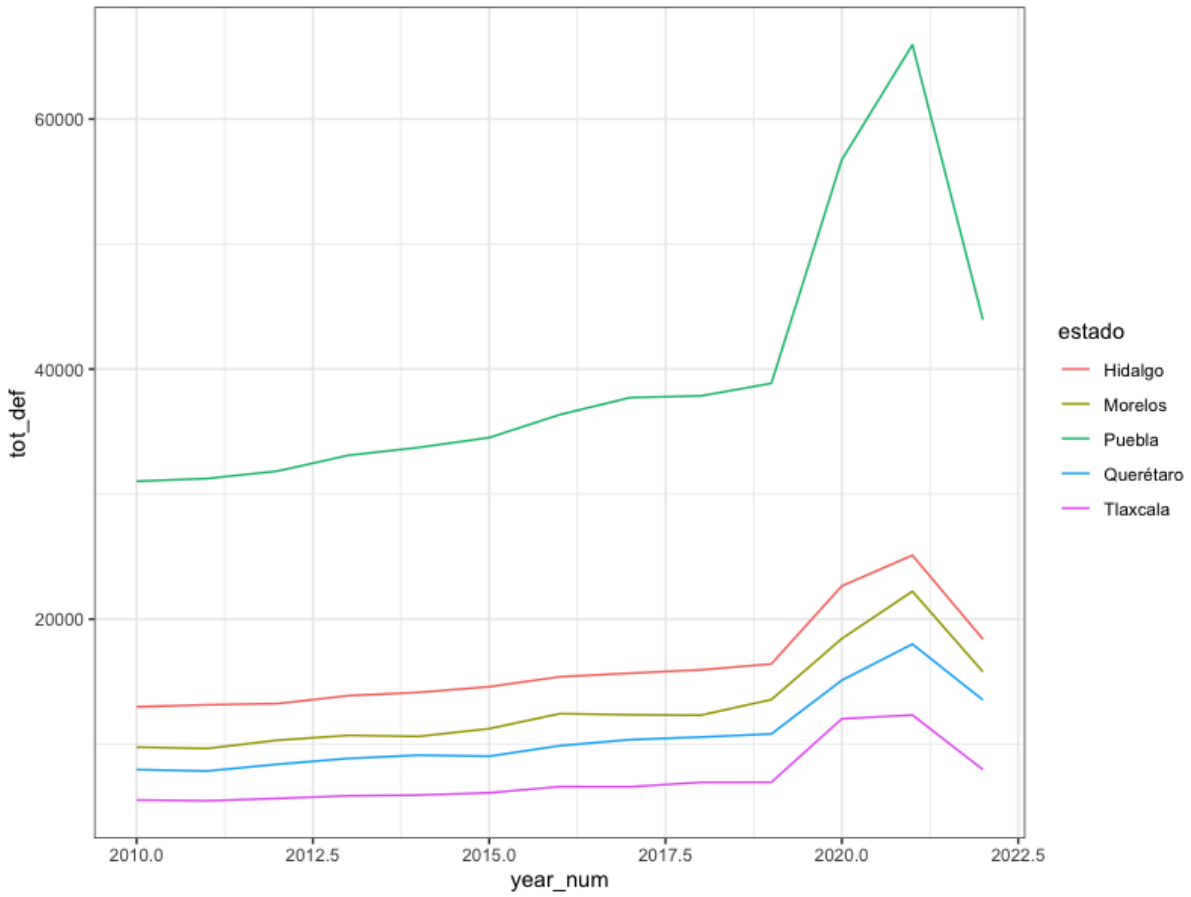


```
defun_l %>%
    filter(!cve_edo %in% c("09", "15")) %>%
    ggplot( )+
    geom_line(aes(year_num, tot_def)) +
    theme_bw()
```

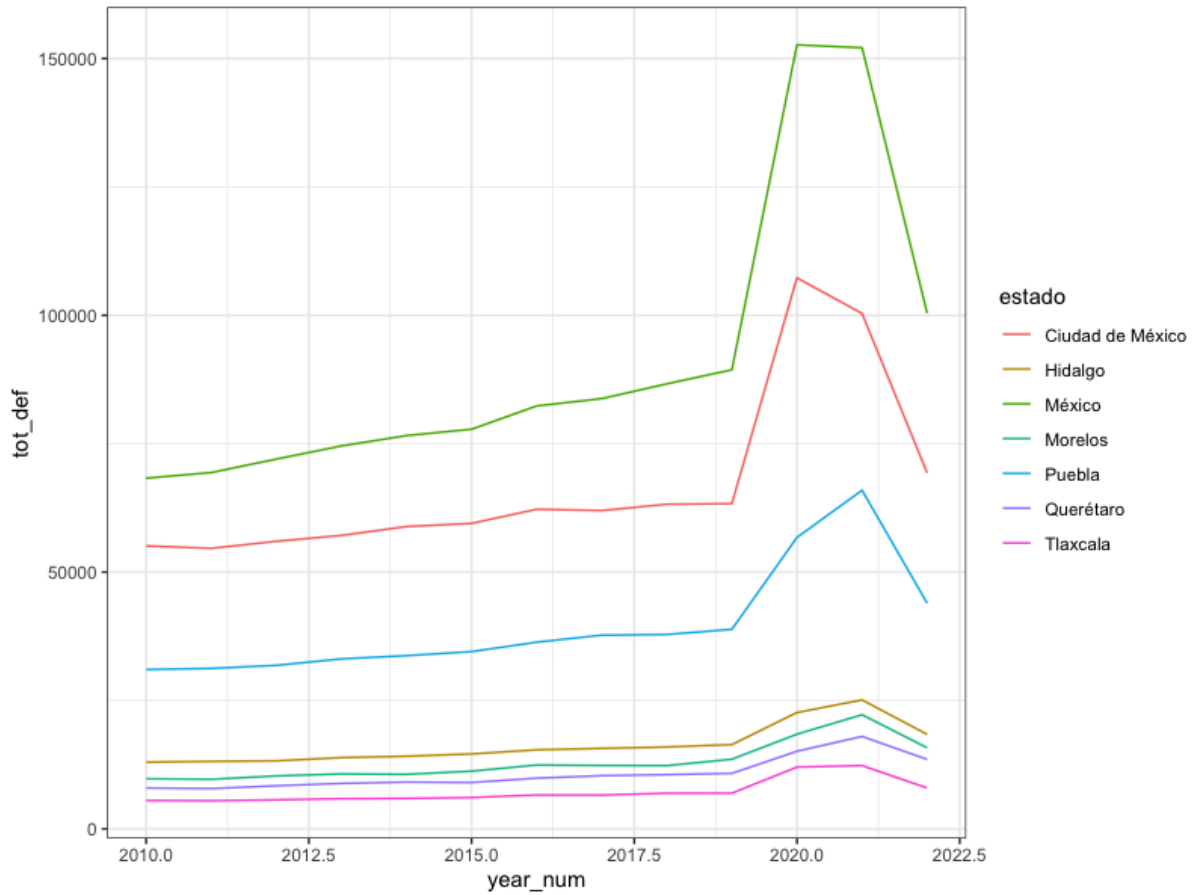
- No es necesario tener la malla como un objeto para graficar.
- Los datos a graficar pueden venir de un preproceso de datos y terminar con el gráfico.



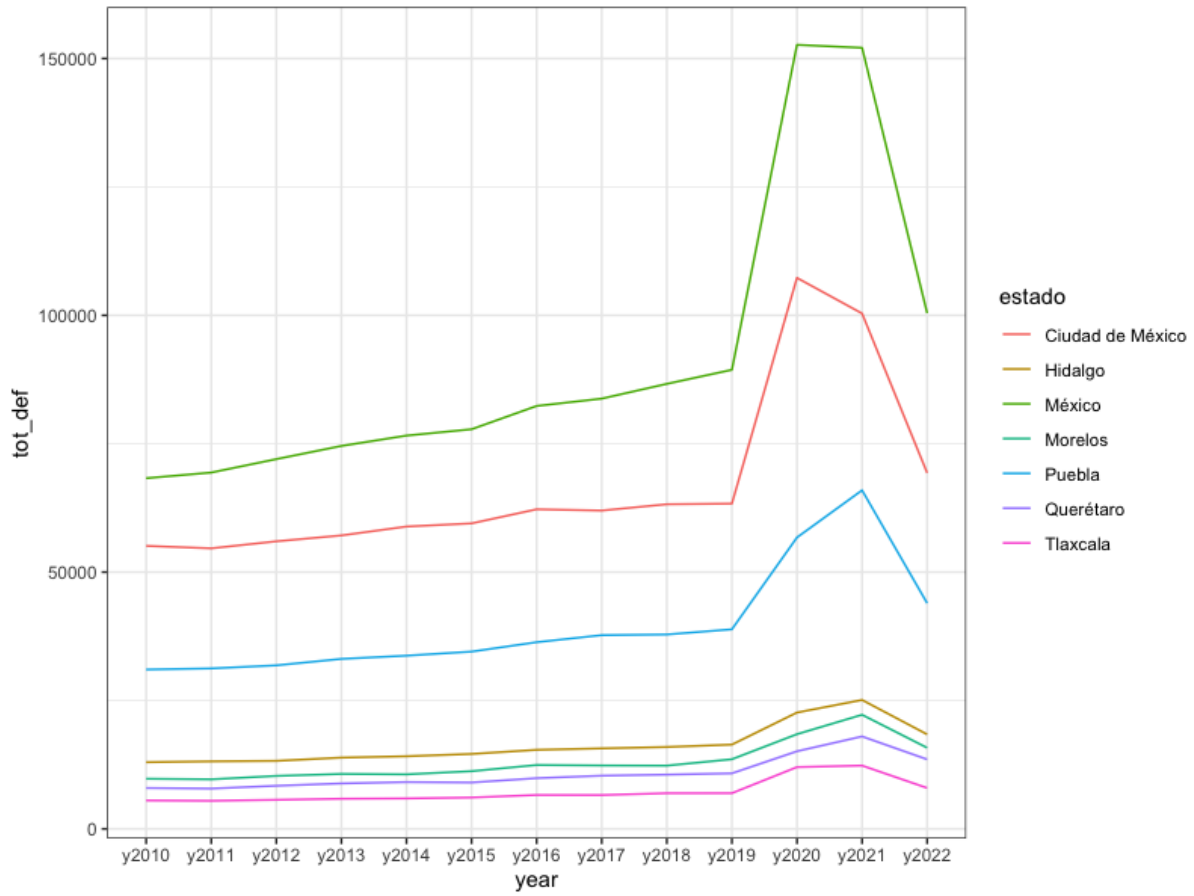
```
defun_l %>%
    filter(!cve_edo %in% c("09", "15")) %>%
    ggplot( )+
    geom_line(aes(year_num, tot_def, color = estado)) +
    theme_bw()
```



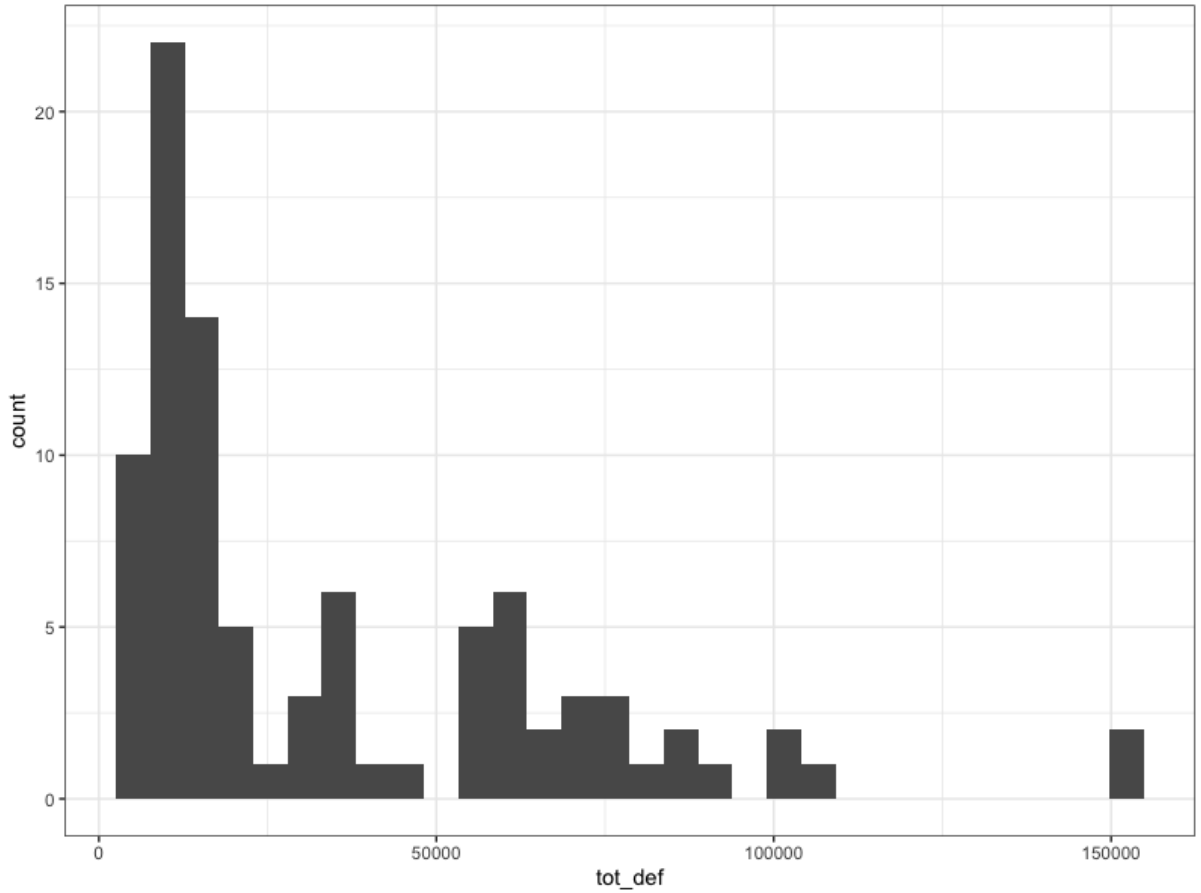
```
ggplot(defun_l)+
    geom_line(aes(year_num, tot_def, color = estado)) +
    theme_bw()
```



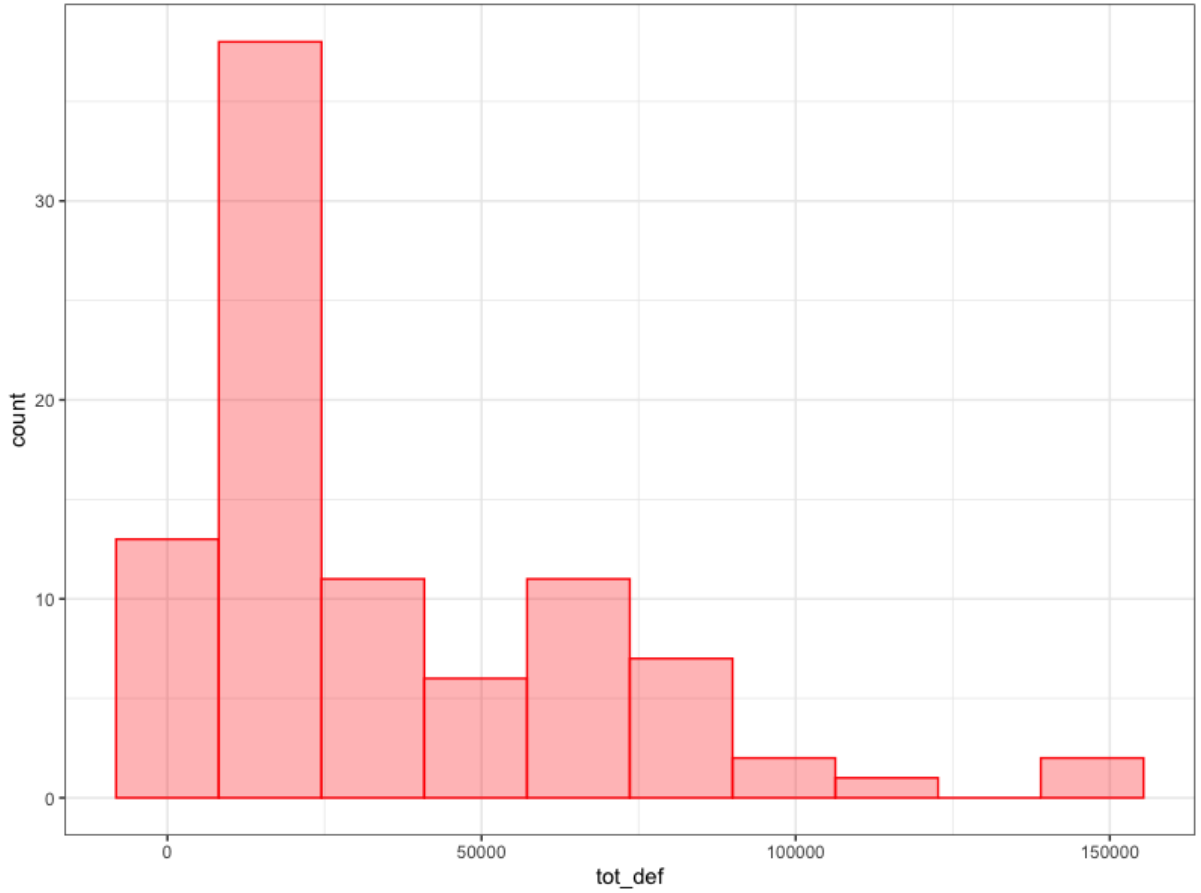
```
ggplot(defun_l)+
    geom_line(aes(year, tot_def, color = estado, group = estado)) +
    theme_bw()
```



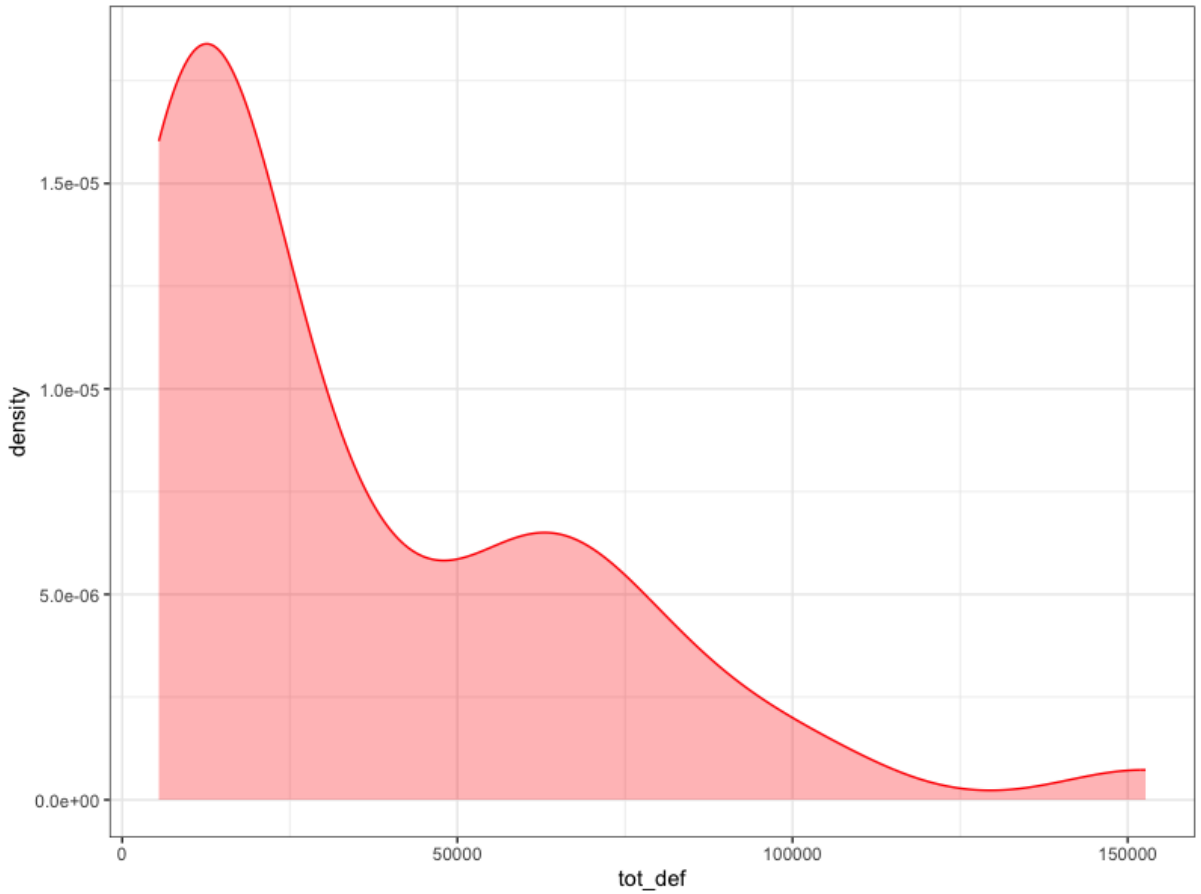
```
ggplot(defun_l) +
    geom_histogram(aes(tot_def)) +
    theme_bw()
```



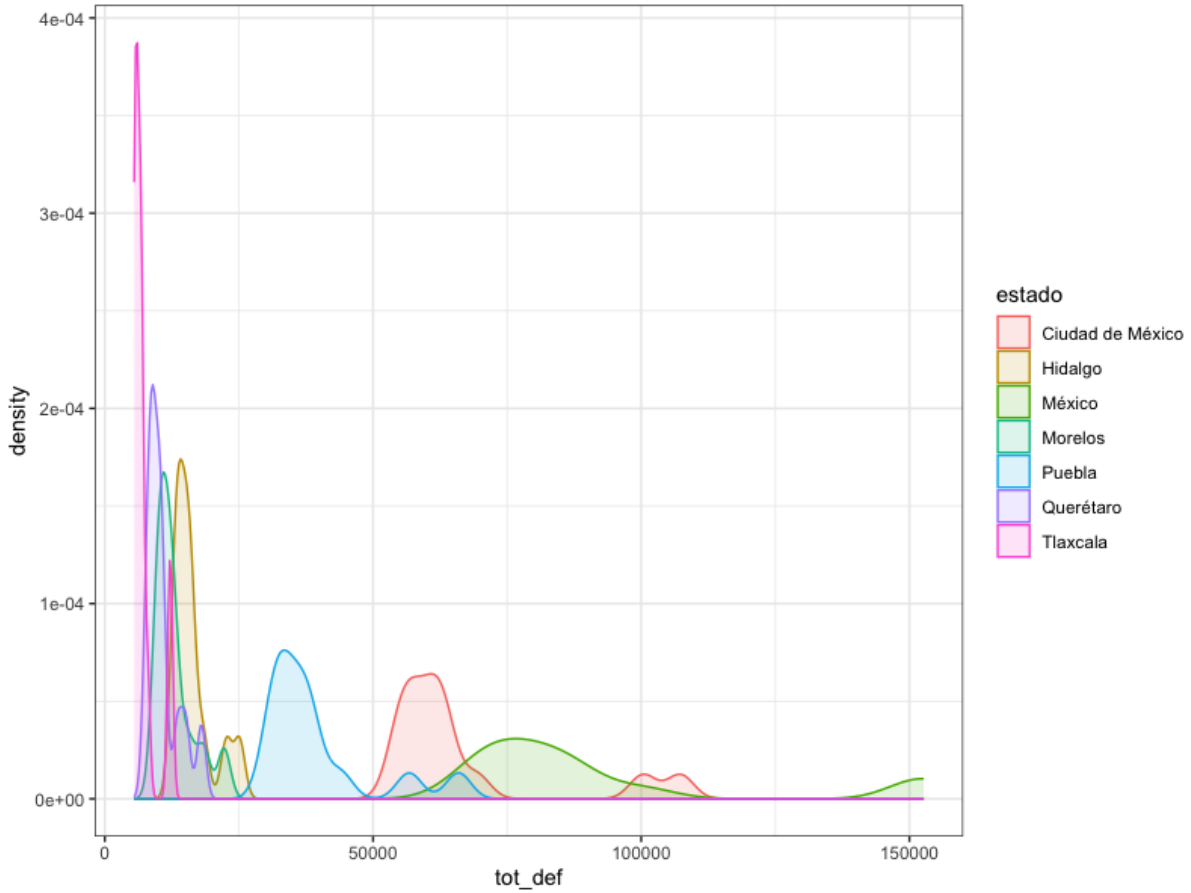
```
ggplot(defun_l) +
    geom_histogram(aes(tot_def), color = "red", fill = alpha("red", 0.3), bins = 10) +
    theme_bw()
```



```
ggplot(defun_l) +
    geom_density(aes(tot_def), color = "red", fill = alpha("red", 0.3)) +
    theme_bw()
```

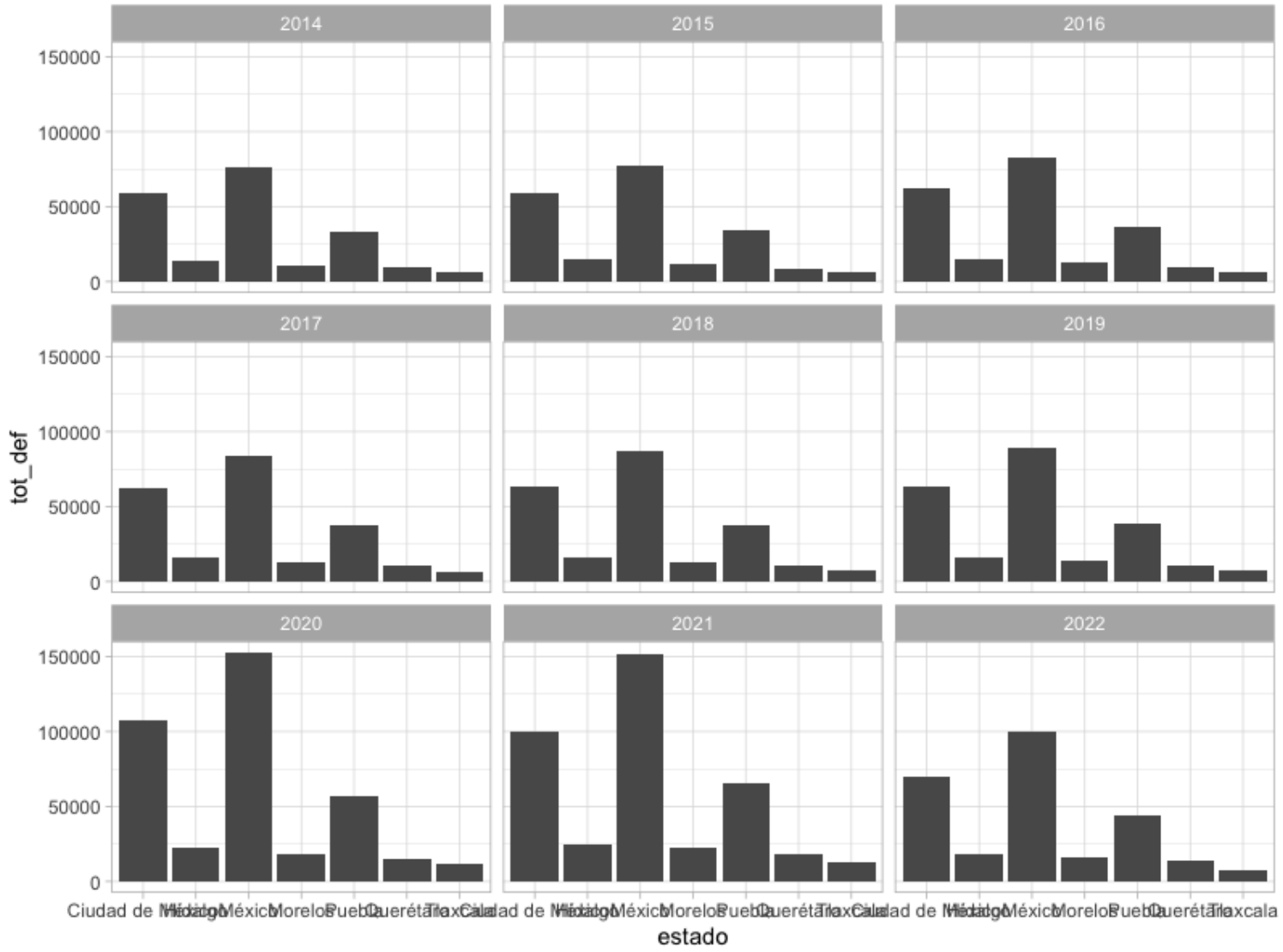


```
ggplot(defun_l) +
    geom_density(aes(tot_def, color = estado, fill = estado), alpha = 0.15) +
    theme_bw()
```



INSP/ESPM
Escuela de Salud Pública de México

```
defun_l %>%
  filter(year_num > 2013) %>%
  ggplot() +
  geom_col(aes(estado, tot_def)) +
  facet_wrap(vars(year_num)) +
  theme_light()
```



Su turno...

- Importe la malla pm25_zmvm_2019.rds y guarde el objeto como pm25_w.
- Transforme a formato long, guarde el resultado como pm25_l.
- Genere un gráfico boxplot.
- Genere un gráfico que muestre la tendencia de la concentración de $PM_{2.5}$ por estación de monitoreo.
- Genere un gráfico que muestre la tendencia de la concentración promedio de $PM_{2.5}$ de todas las estaciones.