



# Visualización de datos geográficos usando software libre

usando R y Leaflet

Federico Aguirre  
[@FAguirreCardiel](#)



# Federico Aguirre

Ingeniero de datos.



Experiencia en arquitecturas de datos: RDBMS, Data Warehouses,  
NOSQL, Data lakes y data pipelines.

Business Intelligence.

Desarrollo de software: Backend y Frontend.

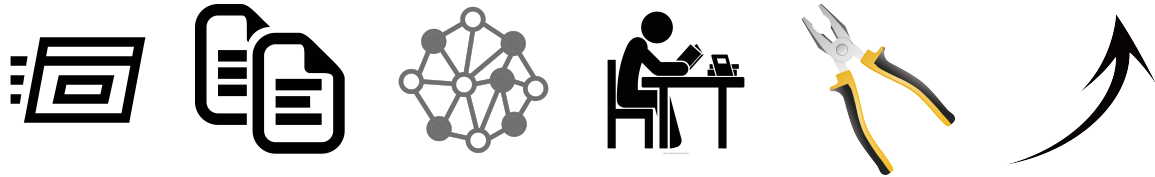
[federico.aguirre.cardiel@gmail.com](mailto:federico.aguirre.cardiel@gmail.com)  
<https://github.com/FedericoAguirre/r-cp-mx>  
@FAguirreCardiel

# Software libre

Es el software que respeta la libertad de los usuarios y la comunidad.

El usuario tiene la libertad de:

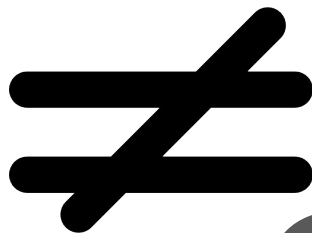
- Ejecutar
- Copiar
- Distribuir
- Estudiar
- Modificar
- Mejorar





**Software libre**

**LIBRE**



**GRATUITO**

A short horizontal bar with a teal segment on the left and an orange segment on the right.

## Lenguaje R

R es un lenguaje y ambiente para gráficos y cómputo estadístico.

Es un proyecto GNU que fue desarrollado en los laboratorios Bell por John Chambers y sus colegas.

R provee una gran variedad de herramientas estadísticas y gráficas y es extensible.



## Lenguaje R y Leaflet

R utiliza distintas librerías o paquetes para poder realizar análisis estadístico.

Los 5 paquetes más usados son:

1. **dplyr**
  2. devtools
  3. foreign
  4. cluster
  5. **ggplot2**
- Leaflet





# Transformar datos en visualizaciones geográficas

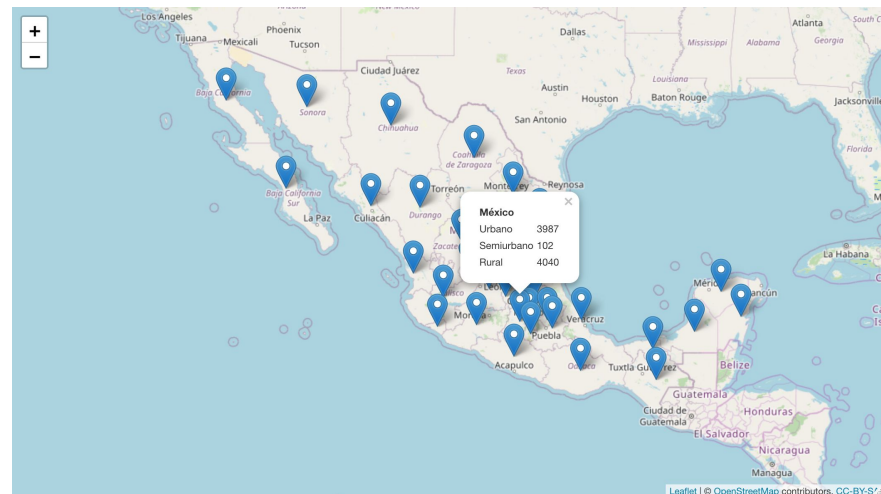
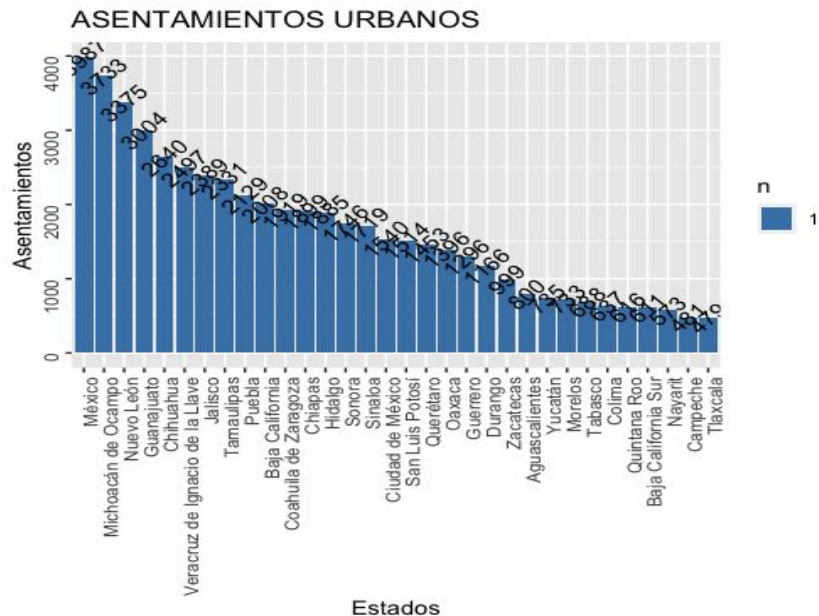
```
→ cp-mx git:(master) ✗ cat CPdescarga.txt | wc -l
143674
```

```
→ cp-mx git:(master) ✗ head -n 15 CPdescarga.txt
```

El Cat?logo Nacional de C?digos Postales, es elaborado por Correos de M?xico y se proporciona en forma gratuita para uso particular, no estando permitida su comercializaci?n, total o parcial, ni su distribuci?n a terceros bajo ning?n concepto.

```
d_codigo|d_asenta|d_tipo_asenta|D_mnpio|d_estado|d_ciudad|d_CP|c_estado|c_oficina|c_CP|c_tipo_asenta|c_mnpio|id_asenta_cpcons|d_zona|c_cve_ciudad
01000|San ?ngel|Colonia|?lvaro Obreg?n|Ciudad de M?xico|Ciudad de M?xico|01001|09|01001||09|010|0001|Urbano|01
01010|Los Alpes|Colonia|?lvaro Obreg?n|Ciudad de M?xico|Ciudad de M?xico|01001|09|01001||09|010|0005|Urbano|01
01020|Guadalupe Inn|Colonia|?lvaro Obreg?n|Ciudad de M?xico|Ciudad de M?xico|01001|09|01001||09|010|0006|Urbano|01
01030|Florida|Colonia|?lvaro Obreg?n|Ciudad de M?xico|Ciudad de M?xico|01001|09|01001||09|010|0010|Urbano|01
01030|Axotla|Pueblo|?lvaro Obreg?n|Ciudad de M?xico|Ciudad de M?xico|01001|09|01001||28|010|0009|Urbano|01
01040|Campestre|Colonia|?lvaro Obreg?n|Ciudad de M?xico|Ciudad de M?xico|01001|09|01001||09|010|0012|Urbano|01
01049|Tlacopac|Pueblo|?lvaro Obreg?n|Ciudad de M?xico|Ciudad de M?xico|01001|09|01001||28|010|0014|Urbano|01
01050|Ex-Hacienda de Guadalupe Chimalistac|Colonia|?lvaro Obreg?n|Ciudad de M?xico|Ciudad de M?xico|01001|09|01001||09|010|0016|Urbano|01
01060|Altavista|Colonia|?lvaro Obreg?n|Ciudad de M?xico|Ciudad de M?xico|01001|09|01001||09|010|0017|Urbano|01
01060|San ?ngel Inn|Colonia|?lvaro Obreg?n|Ciudad de M?xico|Ciudad de M?xico|01001|09|01001||09|010|0018|Urbano|01
01070|Chimalistac|Colonia|?lvaro Obreg?n|Ciudad de M?xico|Ciudad de M?xico|01001|09|01001||09|010|0019|Urbano|01
01080|Progreso Tizapan|Colonia|?lvaro Obreg?n|Ciudad de M?xico|Ciudad de M?xico|01001|09|01001||09|010|0022|Urbano|01
01089|Ermita Tizapan|Colonia|?lvaro Obreg?n|Ciudad de M?xico|Ciudad de M?xico|01001|09|01001||09|010|0024|Urbano|01
```

# Transformar datos en visualizaciones





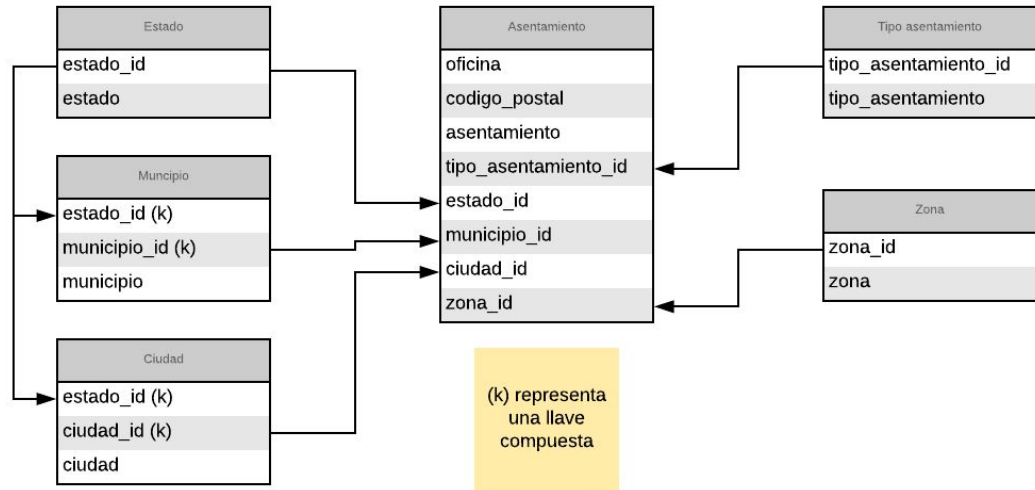
# Manejo y mapeo de datos

d\_codigo  
d\_asenta  
d\_tipo\_asenta  
D\_mnpio  
d\_estado  
d\_ciudad  
d\_CP  
c\_estado  
c\_oficina  
c\_CP  
c\_tipo\_asenta  
c\_mnpio  
id\_asenta\_cpcons  
d\_zona  
c\_cve\_ciudad



## Asentamientos MX ER

Federico Aguirre Cardiel | November 18, 2019





## Funciones usadas de R.

1. **library.** Se utiliza para cargar paquetes que realizan operaciones específicas.
2. **setwd.** Define el directorio de trabajo raíz.
3. **read.csv.** Lee datos (de texto) provenientes de archivos separados por comas (comma separated values).
4. **write.csv.** Escribe datos de matrices o dataframes (tablas) de R en archivos separados por comas (comma separated values).
5. **Funciones propias.** Son funciones que el analista desarrolla por su cuenta para facilitar los cálculos o legibilidad de sus programas.



## Funciones usadas de dplyr.

1. `select.`
2. `distinct.`
3. `filter.`
4. `mutate.`
5. `crossing.`
6. `left_join.`
7. `group_by.`
8. `summarise.`
9. `spread.`
10. `%>%`

[dplyr cheat sheet](#)



## Código de ejemplo R y dplyr.

```
library(dplyr)

setwd("/Users/federicoaguirre/Documents/proyectos/r-cp-mx/")

cp_raw_df <- read.csv("./cp-mx/CPdescarga.txt" , header = TRUE, sep
= "|",
  skip = 1, colClasses = "character", na.strings = "NA", encoding =
"latin1")

zona_df <- select(cp_raw_df, zona = d_zona) %>% distinct()
zona_df <- mutate(zona_df, zona_id = substring(zona, 0, 1)) %>%
  select(zona_id, zona)
write.csv(zona_df, file = "./cp-db/zona.csv", row.names = FALSE)
```



## Código de ejemplo R y dplyr.

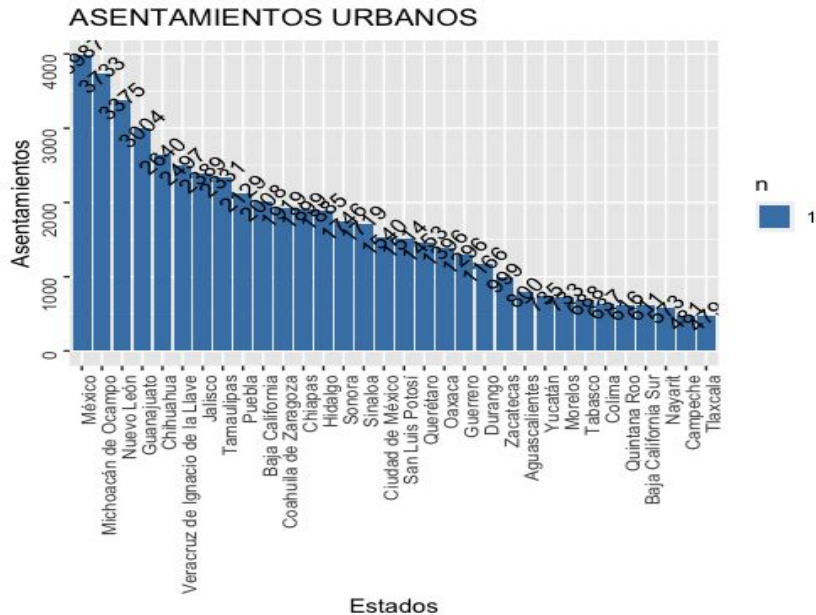
```
conteo_asentamiento_df <- crossing(geoposicion_df, zona_df) %>%  
  left_join(asentamiento_df, by = c("estado_id", "zona_id"), copy =  
FALSE, keep = TRUE) %>%  
  mutate(tiene_asentamiento = ifelse(is.na(asentamiento), 0, 1 )) %>%  
  select(estado_id, estado, latitud, longitud, zona_id, zona,  
tiene_asentamiento) %>%  
  group_by(estado_id, estado, latitud, longitud, zona_id, zona) %>%  
  summarise(asentamientos = sum(tiene_asentamiento))
```

# Funciones usadas de ggplot2.

1. `ggplot.`
2. `aes.`
3. `labs.`
4. `xlab`
5. `ylab`
6. `geombar.`
7. `geom_text.`
8. `theme.`

<https://www.r-graph-gallery.com/>

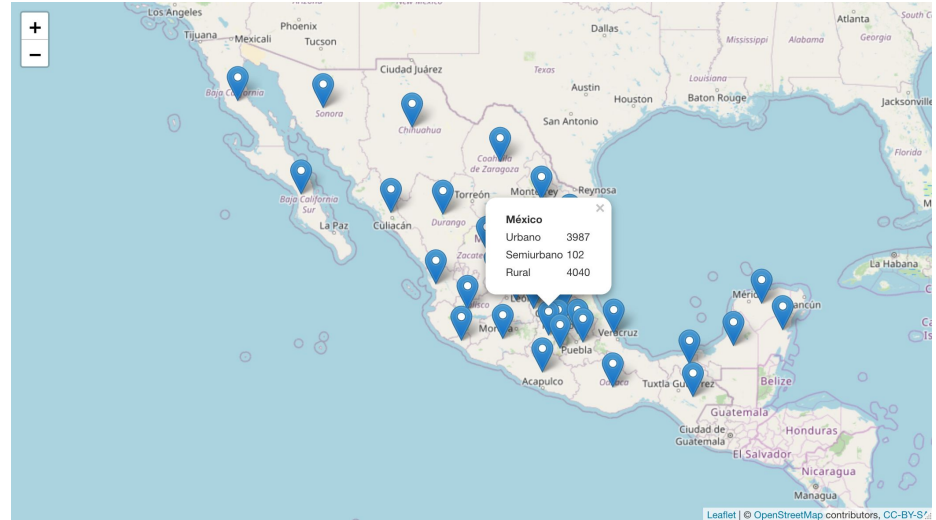
[ggplot2 cheat sheet](#)



# Funciones usadas de leaflet.

1. `leaflet.`
2. `fitBounds.`
3. `addTiles.`
4. `addMarkers`
5. `%>%`

[https://leafletjs.com/  
leaflet cheatsheet](https://leafletjs.com/leaflet-cheatsheet)



  
**Gracias totales.**

