

# Proyecto

2022-05-10

## R Markdown

Cargamos los datos:

```
info <- read.csv('data/datosfinales.csv')
```

## Planteamiento del problema

La Ciudad de México es una de las 10 ciudades más grandes del mundo por población con habitantes al 2020, sin contar la zona metropolitana que incluye zonas del Estado de México e Hidalgo. Una población de este tamaño exige un sistema de transporte público masivo de alta frecuencia, volumen y disponibilidad. El sistema de transporte público unificado en la Ciudad de México es en nuestra opinión uno relativamente bien planeado y accesible. Sin embargo, como personas que no vivimos en la periferia de la zona metropolitana nuestras opiniones pueden estar sesgadas.

El objetivo de esta investigación es cuantificar el nivel de acceso de la población de distintas alcaldías de la ciudad a los diversos medios de transporte público masivo. Como transporte público masivo estamos tomando en cuenta los siguientes servicios de transporte unificado que ofrece la ciudad:

- Metro
- Metrobus
- Tren Ligero
- Cablebus

Elegimos concentrarnos en estos servicios por las siguientes características:

1. Frecuencia. La frecuencia con la que pasan nuevos convoyes debe ser relativamente alta. Por ejemplo, en horas pico pasan convoyes nuevos de metro y metrobus en pocos minutos.
2. Volumen. Nos concentramos en transportes de alto volumen, excluyendo peseros y microbuses.
3. Unificado. Nos concentramos en el sistema de transporte unificado coordinado por el gobierno de la Ciudad de México.

Además de hacer una exploración el objetivo de esta investigación es determinar que tan bien distribuido está el transporte público en la ciudad. Como habitantes de la CDMX tenemos la sospecha de que el transporte público está muy centralizado en la zona del centro histórico. Es decir, sospechamos que el sistema de transporte público privilegia a las personas que viven en las delegaciones como Benito Juárez, Cuauhtémoc, etc... que no son necesariamente las delegaciones con las poblaciones más altas.

Estas relaciones las exploraremos mediante diversas técnicas cubiertas en el curso. Primero que nada procedemos con análisis exploratorio para empezar a ganar intuición sobre el conjunto de datos. Más tarde aplicamos técnicas estadísticas para construir algo como un “índice de conectividad”. Exploramos cómo se relaciona este índice con variables de interés como: centralidad medida en distancia a la zona del centro histórico, población, y otros factores.

## Análisis exploratorio

```
info |> head()
```

```
##      AÑO      ALCALDIA POBLACION MEAN_DIST EST_TOTAL ZOC_DIST
## 1 1969 Azcapotzalco    527857 2019.0774      0 2028.311
## 2 1970 Azcapotzalco    534554 1050.1630      0 2028.311
## 3 1971 Azcapotzalco    541251 1050.1630      0 2028.311
## 4 1972 Azcapotzalco    547948  958.8664      0 2028.311
## 5 1973 Azcapotzalco    554645  958.8664      0 2028.311
## 6 1974 Azcapotzalco    561342  958.8664      0 2028.311
```

```
summary(info)
```

```
##      AÑO      ALCALDIA      POBLACION      MEAN_DIST
## Min.   :1969   Length:848   Min.    : 30700   Min.    : 140.8
## 1st Qu.:1982   Class :character 1st Qu.: 257079   1st Qu.: 382.5
## Median :1995   Mode  :character  Median : 443768   Median : 942.3
## Mean   :1995                      Mean   : 532712   Mean   : 2252.4
## 3rd Qu.:2008                      3rd Qu.: 639251   3rd Qu.: 3051.8
## Max.   :2021                      Max.    :1837010   Max.    :12926.4
##                                     NA's    :53
##      EST_TOTAL      ZOC_DIST
## Min.   : 0.00   Min.    : 2028
## 1st Qu.: 0.00   1st Qu.: 3641
## Median : 8.00   Median : 5831
## Mean   : 13.46   Mean    : 7366
## 3rd Qu.: 21.00   3rd Qu.:12048
## Max.   :118.00   Max.    :17697
## NA's    :159     NA's     :53
```

Hay 16 alcaldías. Cada alcaldía contiene datos de población y conectividad con el transporte público desde el año 1969 hasta el 2021.

Para el análisis que vamos a llevar a cabo recabamos información de diversas fuentes para indicadores de movilidad para las alcaldías de la CDMX y cómo han evolucionado desde principios de la década de los 70.

Tenemos información para 16 alcaldías, las cuales son:

```
unique(info$ALCALDIA) # |> pprint()
```

```
## [1] "Azcapotzalco"      "Coyoacán"          "Gustavo A. Madero"
## [4] "Iztacalco"         "Iztapalapa"        "Tlalpan"
## [7] "Tláhuac"           "Xochimilco"        "Benito Juárez"
## [10] "Cuauhtémoc"        "Miguel Hidalgo"    "Venustiano Carranza"
## [13] "Cuajimalpa"        "Magdalena Contreras" "Milpa Alta"
## [16] "Álvaro Obregón"
```

Tenemos información para los siguientes años:

```
unique(info$AÑO)
```

```
## [1] 1969 1970 1971 1972 1973 1974 1975 1976 1977 1978 1979 1980 1981 1982 1983
## [16] 1984 1985 1986 1987 1988 1989 1990 1991 1992 1993 1994 1995 1996 1997 1998
## [31] 1999 2000 2001 2002 2003 2004 2005 2006 2007 2008 2009 2010 2011 2012 2013
## [46] 2014 2015 2016 2017 2018 2019 2020 2021
```

Primero que nada, vemos algunas estadísticas de resumen de los datos.

```
summary(info)
```

```
##      AÑO      ALCALDIA      POBLACION      MEAN_DIST
## Min.   :1969   Length:848   Min.    : 30700   Min.    : 140.8
## 1st Qu.:1982   Class :character   1st Qu.: 257079   1st Qu.: 382.5
## Median :1995   Mode  :character   Median : 443768   Median : 942.3
## Mean   :1995                      Mean   : 532712   Mean   : 2252.4
## 3rd Qu.:2008                      3rd Qu.: 639251   3rd Qu.: 3051.8
## Max.   :2021                      Max.    :1837010   Max.    :12926.4
##                                     NA's    :53
## EST_TOTAL      ZOC_DIST
## Min.   : 0.00   Min.    : 2028
## 1st Qu.: 0.00   1st Qu.: 3641
## Median : 8.00   Median : 5831
## Mean   : 13.46   Mean   : 7366
## 3rd Qu.: 21.00   3rd Qu.:12048
## Max.   :118.00   Max.    :17697
## NA's    :159     NA's     :53
```

## Descripción de las variables de interés

Las variables son las siguientes:

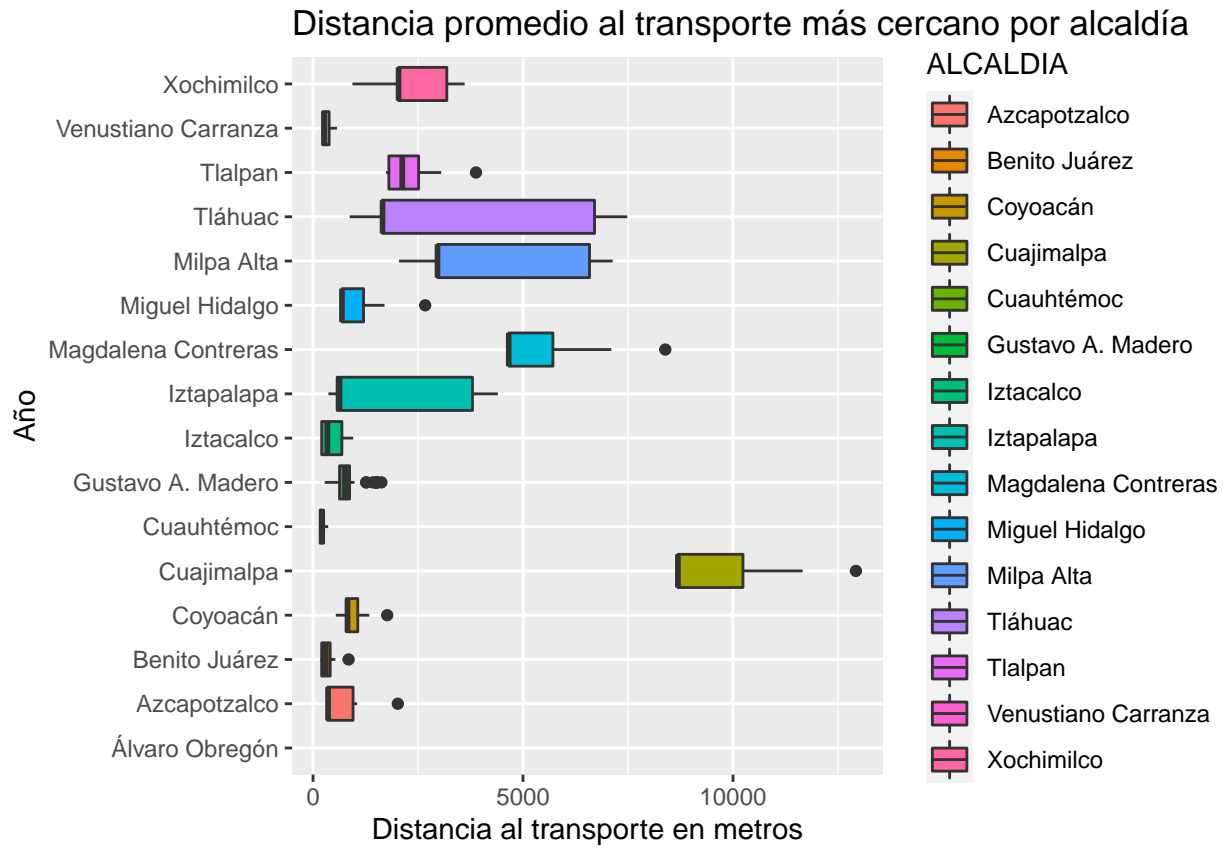
- AÑO: El año en el cual están medidas las variables.
- ALCALDIA: La demarcación territorial de delegación o Alcaldía al que corresponden los datos.
- POBLACIÓN: La población para cada alcaldía en el año dado.
- MEAN\_DIST: La distancia promedio de todas las zonas marcadas como residenciales en la encuesta de uso de suelo a su estación de transporte público masivo más cercano medida en metros.
- EST\_TOTAL: Número total de estaciones de transporte público masivo en la alcaldía al año marcado.
- ZOC\_DIST: Distancia promedio de las zonas residenciales al zócalo de la ciudad.

Las variables fueron construidas a partir de diferentes conjuntos de datos abiertos al público. No encontramos una base de datos que tuviera lista para usarse toda la información que era necesaria para el análisis, menos aún como función del tiempo. En las siguientes secciones ahondamos en algunos detalles técnicos de cómo se obtuvieron, limpiaron, y trabajaron datos faltantes.

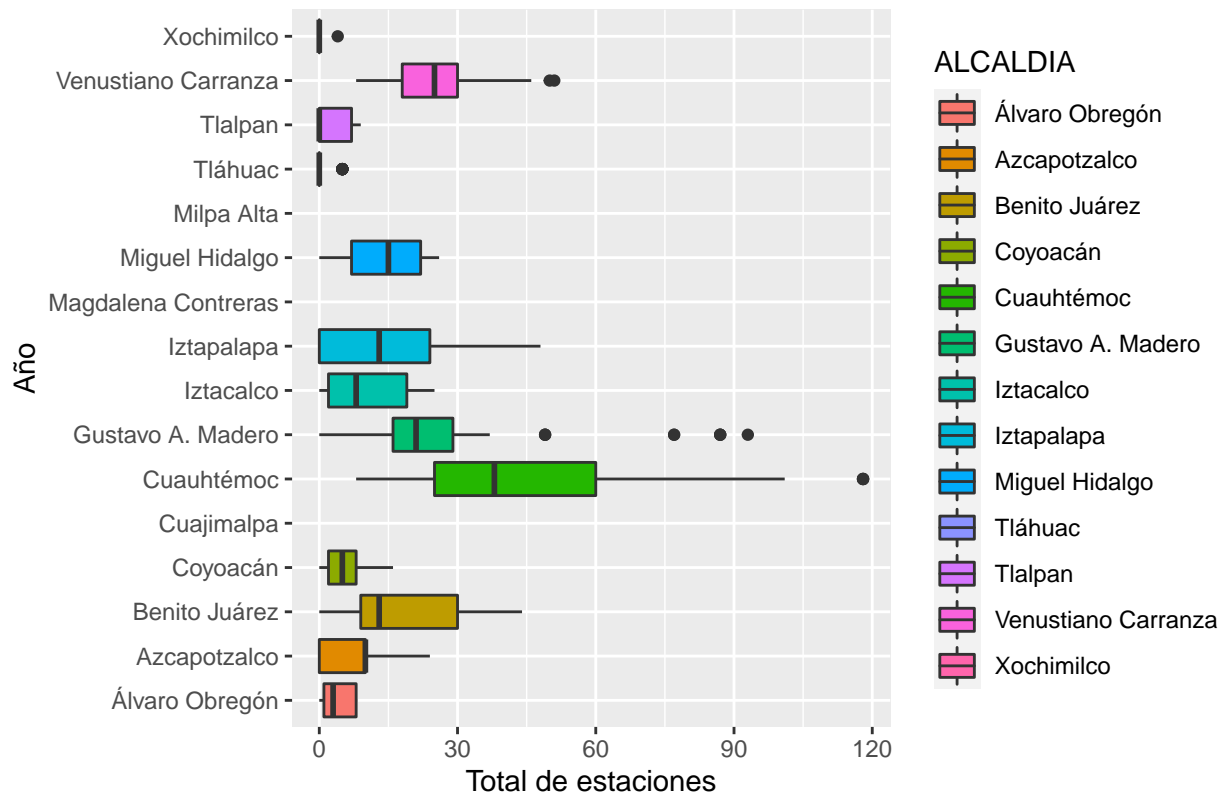
## Número total de estaciones por delegación

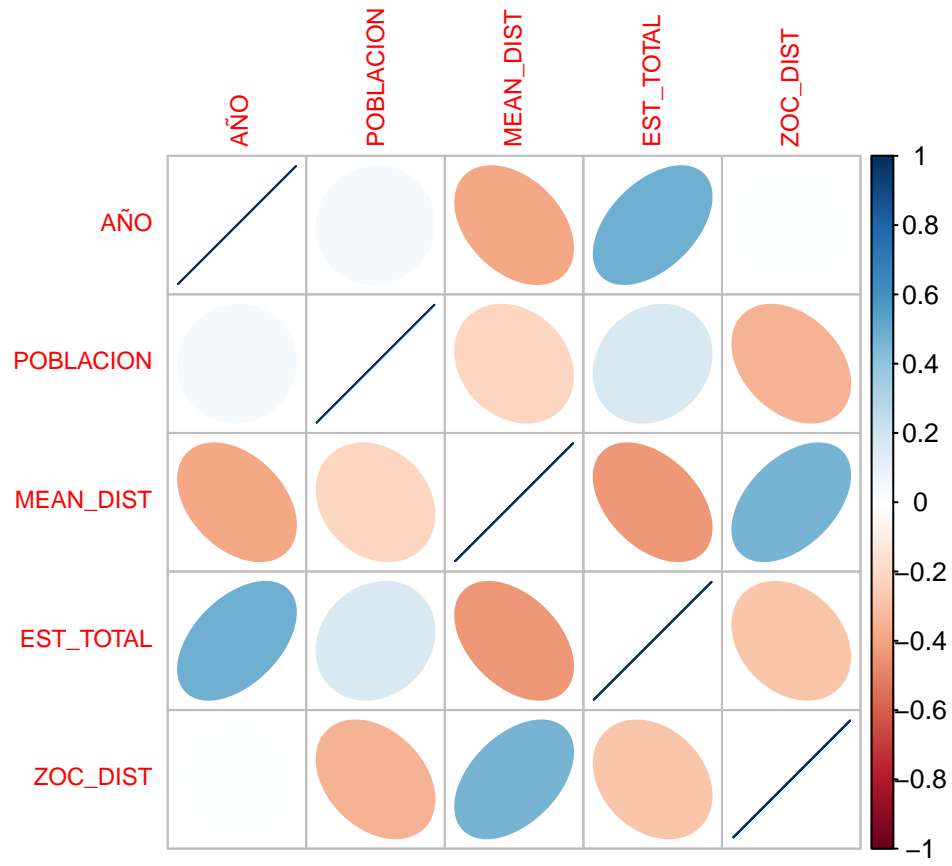
Para encontrar el número de estaciones de transporte público masivo por delegación a un año dado utilizamos los conjuntos de datos [?].

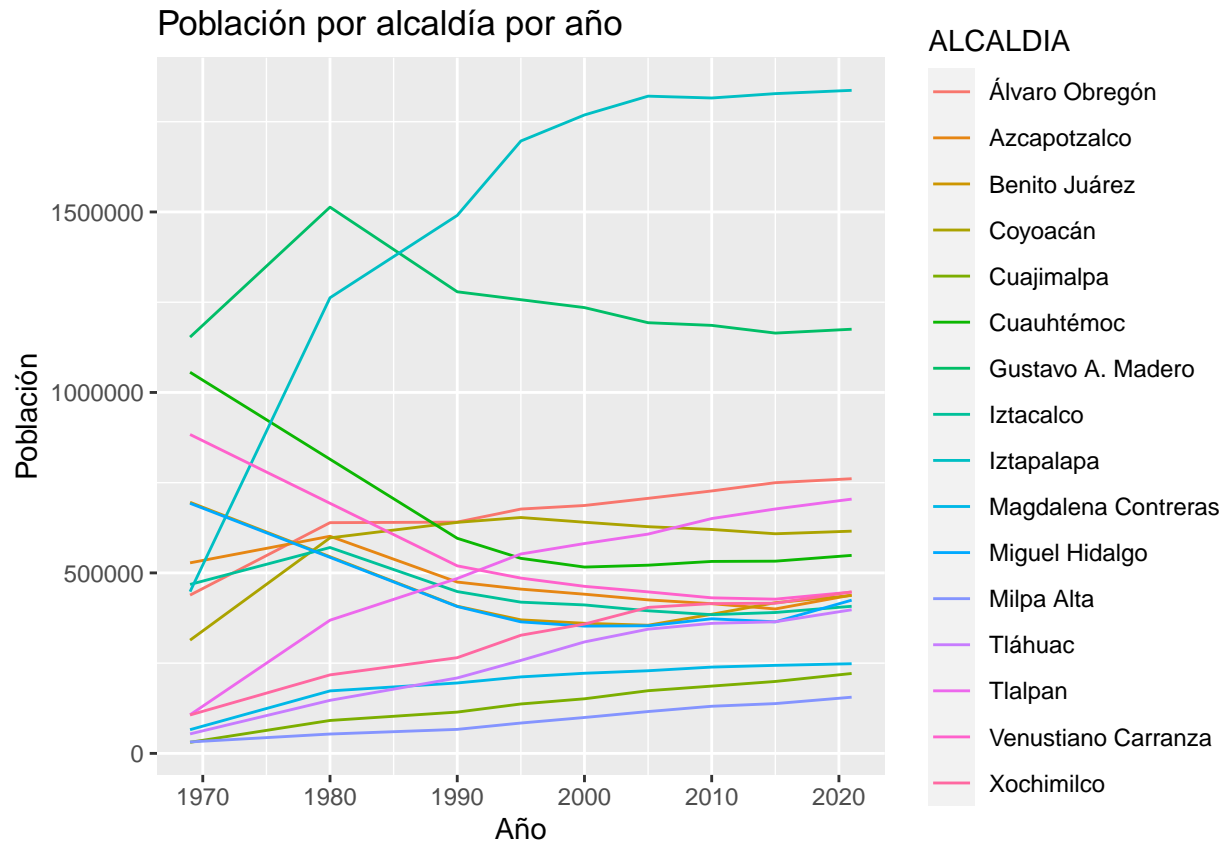
- Podemos ver que las alcaldías de Cuauhtémoc, Benito Juárez y Venustiano Carranza tienen las menores varianzas en distancia media y además las más pequeñas. Por otro lado, Tláhuac y Milpa Alta son los de mayor distancia y desviación.



Total de estaciones por alcaldía a través del tiempo.







**PCA**

**Correlación canónica**

**Analisis Factorial**

**Pruebas de hipótesis**

**Regresión lineal**

**Interpretación, conclusiones, etc...**