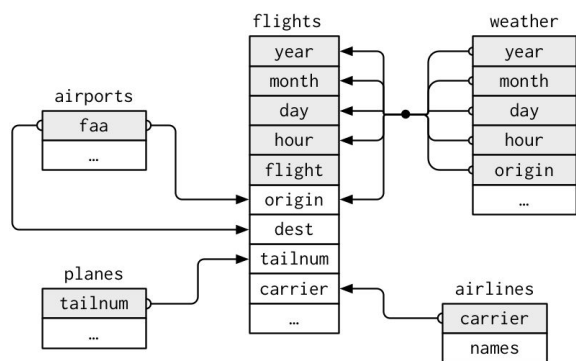


# Retrasos en los vuelos de la ciudad de Nueva York, año 2013

Pablo Peitsch, [preynosopeitsch@estudiantes.unsam.edu.ar](mailto:preynosopeitsch@estudiantes.unsam.edu.ar)

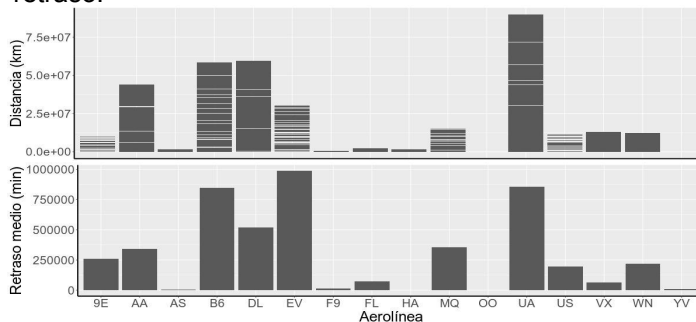
**INTRODUCCIÓN.** En este documento se hace un breve análisis del set de datos de vuelos de ciudad de Nueva York de 2013 (nycflights13), que se encuentra disponible como paquete en los repositorios de R [1]. Los datos se encuentran distribuidos en cinco dataframes, Fig. 1. A las compañías les interesa saber con los datos obtenidos si es posible determinar a qué se deben los retrasos de sus vuelos en los distintos aeropuertos y si es posible evitarlos. Se trabaja principalmente sobre el dataframe “flights” para estudiar los vuelos que presentan retrasos.

En la Fig. 1, se muestra la estructura del set de datos y cómo se relacionan los dataframes.

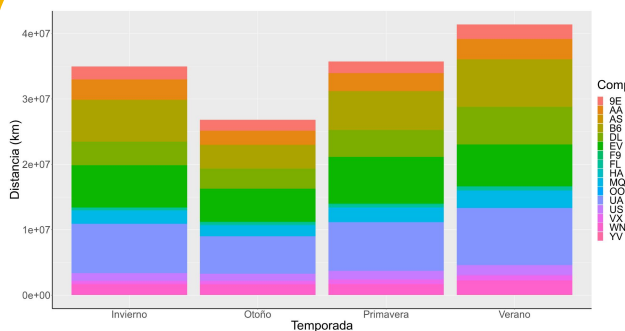


**Figura 1.** Relación entre dataframes, claves foráneas y primarias.

Se grafican las medias de retrasos y la distancia recorrida de los vuelos por las distintas compañías. Se puede ver que los que tienen mayor recorrido presentan mayor retraso.



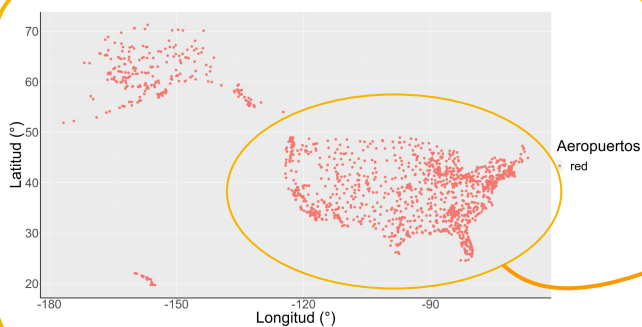
**Figura 2. a.** Retrasos en partida y arribo; **b.** Distancia recorrida, evaluados por aerolínea para todos los aeropuertos.



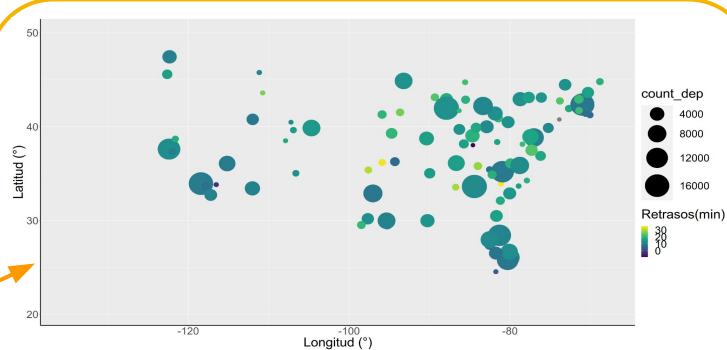
**Figura 3.** Distancia recorrida por los vuelos de las distintas compañías

Se puede apreciar en la Fig. 3 cómo aumenta la distancia recorrida de acuerdo a la temporada. Probablemente esto se deba a que los usuarios utilizan el servicio para vacacionar mayormente en verano y esto aumenta el tráfico.

Para evaluar si los retrasos se originan en determinados aeropuertos, en la Fig. 4, primero se grafican los vuelos en términos de la latitud y longitud registrada. Se puede ver que existe una mayor densidad de puntos, que es donde se encuentran la mayor cantidad de aeropuertos. Por último, se muestran los vuelos con retrasos, donde el tamaño y el color indican el número y la media de retrasos, respectivamente. Si bien no hay una clara distribución espacial, se puede observar que a mayor número de aeropuertos menores retrasos.



**Figura 4.** Distribución espacial aeropuertos EEUU.



**Figura 5.** Distribución espacial de retrasos sobre el mapa de EE.UU.

**CONCLUSIONES.** El análisis realizado mostró que ciertas compañías son propensas a tener mayor retrasos que otras. Principalmente B6, DL, EV y UA. Estas mismas, son aquellas que realizan mayores recorridos de vuelo. A su vez, se distinguen mayores retrasos en ciertas temporadas que en otras, por ejemplo en verano. Una posible solución a esto sería aumentar la frecuencia de vuelos disponibles para evitar retrasos a futuro. Por otro lado, no fue posible determinar con claridad si los retrasos se originan en determinados aeropuertos o zonas geográficas. Es conveniente a futuro revisar otros parámetros que puedan influir en estos retrasos, como ser condiciones meteorológicas.

## REFERENCIAS.

1. <https://rdocumentation.org/packages/nycflights13/versions/1.0.1>
2. <https://github.com/ropensci/rnaturalearth>