

Tarea y Solución

Pregunta 1

Piensa cómo podrías usar la función `arrange()` para colocar todos los valores `NA` al inicio. Pista: puedes la función `is.na()` en lugar de la función `desc()` como argumento de `arrange`.

Solución

```
arrange(flights,!is.na(dep_time))
```

Pregunta 2

Ordena los vuelos de `flights` para encontrar los vuelos más retrasados en la salida. ¿Qué vuelos fueron los que salieron los primeros antes de lo previsto?

Solución

El más retrasado fue:

```
arrange(flights, desc(dep_delay))[1,]
```

El menos retrasado fue:

```
arrange(flights, dep_delay)[1,]
```

Pregunta 3

Ordena los vuelos de `flights` para encontrar los vuelos más rápidos. Usa el concepto de rapidez que consideres.

Solución

```
arrange(flights, desc(distance/air_time))
```

Pregunta 4

¿Qué vuelos tienen los trayectos más largos? Busca en Wikipedia qué dos aeropuertos del dataset alojan los vuelos más largos.

Solución

Vuelos entre el JFK de Nueva York y el HNL, aeropuerto internacional de Honolulu en Hawaii

Pregunta 5

¿Qué vuelos tienen los trayectos más cortos? Busca en Wikipedia qué dos aeropuertos del dataset alojan los vuelos más cortos

Solución

Vuelos entre el EWR, Aeropuerto Internacional Libertad de Newark y LGA, Aeropuerto de La Guardia, ambos situados en el estado de Nueva York

Pregunta 6

Dale al coco para pensar cuántas más maneras posibles de seleccionar los campos `dep_time`, `dep_delay`, `arr_time` y `arr_delay` del dataset de `flights`.

Solución

Puedes hacerlo directamente o bien usando las diferentes funciones que hemos visto en la sección. Algunos ejemplos son:

```
select(flights, dep_time, dep_delay, arr_time, arr_delay)
select(flights, starts_with("dep"), starts_with("arr"))
select(flights, (ends_with("time") | ends_with("delay")) & (starts_with("dep") | starts_with("arr")))
```

Pregunta 7

¿Qué ocurre si pones el nombre de una misma variable varias veces en un `select()`?

Solución

Prueba con la siguiente y verás que solo te lo da una vez.

```
select(flights, distance, distance)
```

Pregunta 8

Investiga el uso de la función `one_of()` de `dplyr`

Solución

Permite añadir las variables en string dentro de un vector. Muy útil si es el resultado de un programa que ha devuelto un array de variables que queremos seleccionar automáticamente.

Pregunta 9

Investiga cómo puede ser útil la función `one_of()` de la pregunta anterior en conjunción con el vector de variables

```
c("year", "month", "day", "dep_delay", "arr_delay")
```

Solución

```
select(flights, one_of(c("year", "month", "day", "dep_delay", "arr_delay")))
```

Pruébalo y verás qué útil es para filtrar como resultado de un subprograma.

Pregunta 10

Intenta averiguar el resultado del siguiente código. Luego, ejecútalo y a ver si el resultado te sorprende.

```
select(flights, contains("time"))
```

Intenta averiguar cómo lo hacen las funciones de ayuda de la función `select` para tratar el caso por defecto y cómo lo puedes cambiar.

Solución

Nos devuelve todas las variables que tienen la palabra “time” en su nombre. Muy útil cuando queremos localizar todo lo que tiene que ver con tiempo.