

Tarea9

Yimmy Eman

2022-07-08

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`.

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

y...	m...	...	dep_time	sched_dep_time	dep_delay	arr_time	sched_arr_time	arr_delay	c
<int>	<int>	<int>	<int>	<int>	<dbl>	<int>	<int>	<dbl>	<
2013	1	2	NA	1545	NA	NA	1910	NA	A
2013	1	2	NA	1601	NA	NA	1735	NA	U
2013	1	3	NA	857	NA	NA	1209	NA	U
2013	1	3	NA	645	NA	NA	952	NA	U
2013	1	4	NA	845	NA	NA	1015	NA	9
2013	1	4	NA	1830	NA	NA	2044	NA	9
2013	1	5	NA	840	NA	NA	1001	NA	9
2013	1	7	NA	820	NA	NA	958	NA	9
2013	1	8	NA	1645	NA	NA	1838	NA	U
2013	1	9	NA	755	NA	NA	1012	NA	9

1-10 of 10,000 rows | 1-10 of 19 columns

Previous 1 2 3 4 5 6 ... 1000 Next

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?

```
arrange(flights, desc(dep_delay))[1,] # Más retrasado
```

y...	m...	...	dep_time	sched_dep_time	dep_delay	arr_time	sched_arr_time	arr_delay	c
<int>	<int>	<int>	<int>	<int>	<dbl>	<int>	<int>	<dbl>	<
2013	1	9	641	900	1301	1242	1530	1272	H

1 row | 1-10 of 19 columns

```
arrange(flights, dep_delay)[1,] # Menos retrasado
```

y...	m...	...	dep_time	sched_dep_time	dep_delay	arr_time	sched_arr_time	arr_delay	c
<int>	<int>	<int>	<int>	<int>	<dbl>	<int>	<int>	<dbl>	<
2013	12	7	2040	2123	-43	40	2352	48	B

1 row | 1-10 of 19 columns

Pregunta 3

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

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

y...	m...	...	dep_time	sched_dep_time	dep_delay	arr_time	sched_arr_time	arr_delay	c
<int>	<int>	<int>	<int>	<int>	<dbl>	<int>	<int>	<dbl>	<
2013	5	25	1709	1700	9	1923	1937	-14	D
2013	7	2	1558	1513	45	1745	1719	26	E
2013	5	13	2040	2025	15	2225	2226	-1	E
2013	3	23	1914	1910	4	2045	2043	2	E
2013	1	12	1559	1600	-1	1849	1917	-28	D
2013	11	17	650	655	-5	1059	1150	-51	D
2013	2	21	2355	2358	-3	412	438	-26	B
2013	11	17	759	800	-1	1212	1255	-43	A
2013	11	16	2003	1925	38	17	36	-19	D
2013	11	16	2349	2359	-10	402	440	-38	B

Pregunta 4

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

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

```
filter(flights, origin == "JFK", dest == "HNL")
```

y...	m...	...	dep_time	sched_dep_time	dep_delay	arr_time	sched_arr_time	arr_delay	c
<int>	<int>	<int>	<int>	<int>	<dbl>	<int>	<int>	<dbl>	<
2013	1	1	857	900	-3	1516	1530	-14	H
2013	1	2	909	900	9	1525	1530	-5	H
2013	1	3	914	900	14	1504	1530	-26	H
2013	1	4	900	900	0	1516	1530	-14	H
2013	1	5	858	900	-2	1519	1530	-11	H
2013	1	6	1019	900	79	1558	1530	28	H
2013	1	7	1042	900	102	1620	1530	50	H
2013	1	8	901	900	1	1504	1530	-26	H
2013	1	9	641	900	1301	1242	1530	1272	H
2013	1	10	859	900	-1	1449	1530	-41	H

Pregunta 5

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

Vuelos entre el EWR, Aeropuerto Internacional Libertad de Newark y LGA

```
filter(flights, origin == "EWR", dest == "LGA")
```

y...	m...	...	dep_time	sched_dep_time	dep_delay	arr_time	sched_arr_time	arr_delay	c
<int>	<int>	<int>	<int>		<dbl>	<int>		<dbl>	<
2013	7	27	NA	106	NA	NA	245	NA	U

1 row | 1-10 of 19 columns

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.

```
select(flights, dep_time, dep_delay, arr_time, arr_delay)
```

dep_time	dep_delay	arr_time	arr_delay
<int>	<dbl>	<int>	<dbl>
517	2	830	11
533	4	850	20
542	2	923	33
544	-1	1004	-18
554	-6	812	-25
554	-4	740	12
555	-5	913	19
557	-3	709	-14
557	-3	838	-8
558	-2	753	8

1-10 of 10,000 rows

Previous 1 2 3 4 5 6 ... 1000 Next

```
select(flights, starts_with("dep") | starts_with("arr"))
```

dep_time	dep_delay	arr_time	arr_delay
<int>	<dbl>	<int>	<dbl>
517	2	830	11
533	4	850	20

542	2	923	33
544	-1	1004	-18
554	-6	812	-25
554	-4	740	12
555	-5	913	19
557	-3	709	-14
557	-3	838	-8
558	-2	753	8

1-10 of 10,000 rows Previous 1 2 3 4 5 6 ... 1000 Next

```
select(flights, contains("dep") | contains("arr"))
```

dep_time <int>	sched_dep_time <int>	dep_delay <dbl>	arr_time <int>	sched_arr_time <int>	arr_delay <dbl>	carrier <chr>
517	515	2	830	819	11	UA
533	529	4	850	830	20	UA
542	540	2	923	850	33	AA
544	545	-1	1004	1022	-18	B6
554	600	-6	812	837	-25	DL
554	558	-4	740	728	12	UA
555	600	-5	913	854	19	B6
557	600	-3	709	723	-14	EV
557	600	-3	838	846	-8	B6
558	600	-2	753	745	8	AA

1-10 of 10,000 rows Previous 1 2 3 4 5 6 ... 1000 Next

Pregunta 7

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

Se selecciona la columna 1 sola vez.

```
select(flights, dep_time, dep_time)
```

	dep_time <int>
	517
	533
	542
	544
	554
	554
	555
	557
	557
	558

1-10 of 10,000 rows

Previous 1 2 3 4 5 6 ... 1000 Next

Pregunta 8

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

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")`

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

year <int>	month <int>	day <int>	dep_delay <dbl>	arr_delay <dbl>
2013	1	1	2	11
2013	1	1	4	20

2013	1	1	2	33
2013	1	1	-1	-18
2013	1	1	-6	-25
2013	1	1	-4	12
2013	1	1	-5	19
2013	1	1	-3	-14
2013	1	1	-3	-8
2013	1	1	-2	8

1-10 of 10,000 rows

Previous 1 2 3 4 5 6 ... 1000 Next

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.

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

dep_time <int>	sched_dep_time <int>	arr_time <int>	sched_arr_time <int>	air_time <dbl>	time_hour <dtm>
517	515	830	819	227	2013-01-01 05:00:00
533	529	850	830	227	2013-01-01 05:00:00
542	540	923	850	160	2013-01-01 05:00:00
544	545	1004	1022	183	2013-01-01 05:00:00
554	600	812	837	116	2013-01-01 06:00:00
554	558	740	728	150	2013-01-01 05:00:00
555	600	913	854	158	2013-01-01 06:00:00
557	600	709	723	53	2013-01-01 06:00:00
557	600	838	846	140	2013-01-01 06:00:00
558	600	753	745	138	2013-01-01 06:00:00

1-10 of 10,000 rows

Previous 1 2 3 4 5 6 ... 1000 Next

Útil para filtrar todo lo que tiene que ver con el tiempo