Licenciatura en
Ciencia de Datos
ECYT_UNSAM

Métricas de resumen

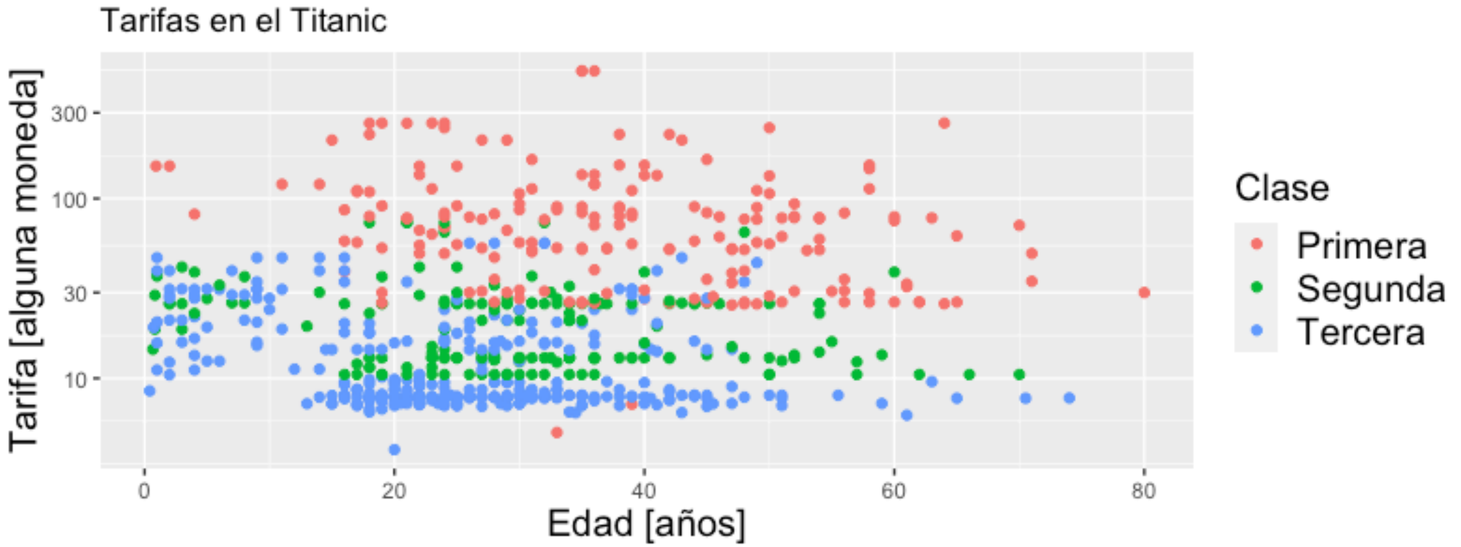
Campus Miguelete, 18 de marzo de 2024

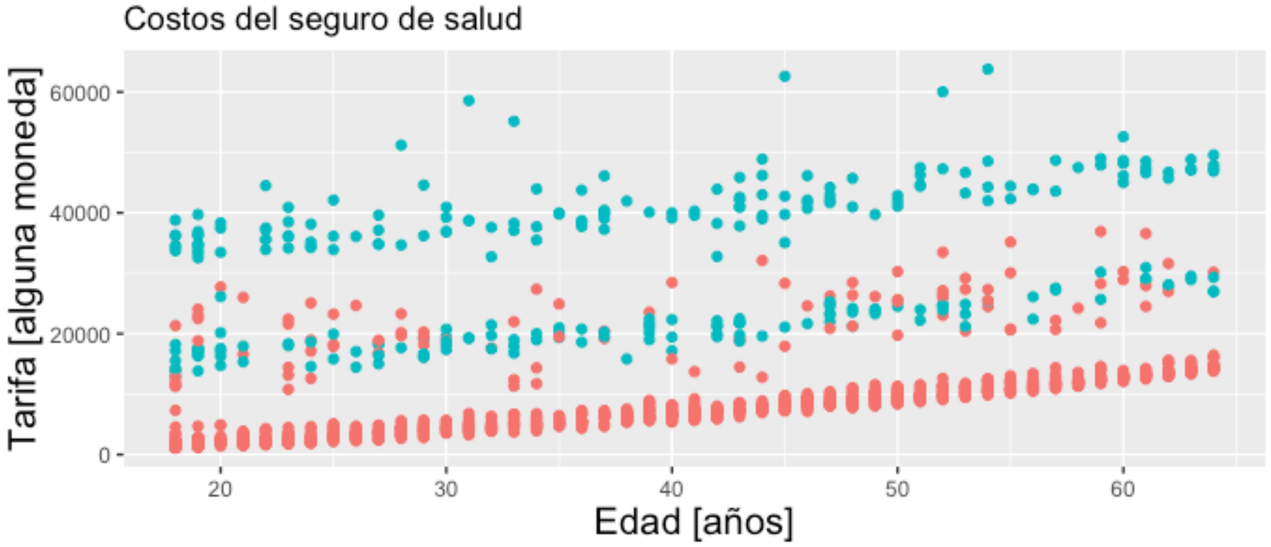
Hasta ahora

- Exploración del dataset
- Manipulación (orden, filtrado, agrupamiento)

Algunas preguntas

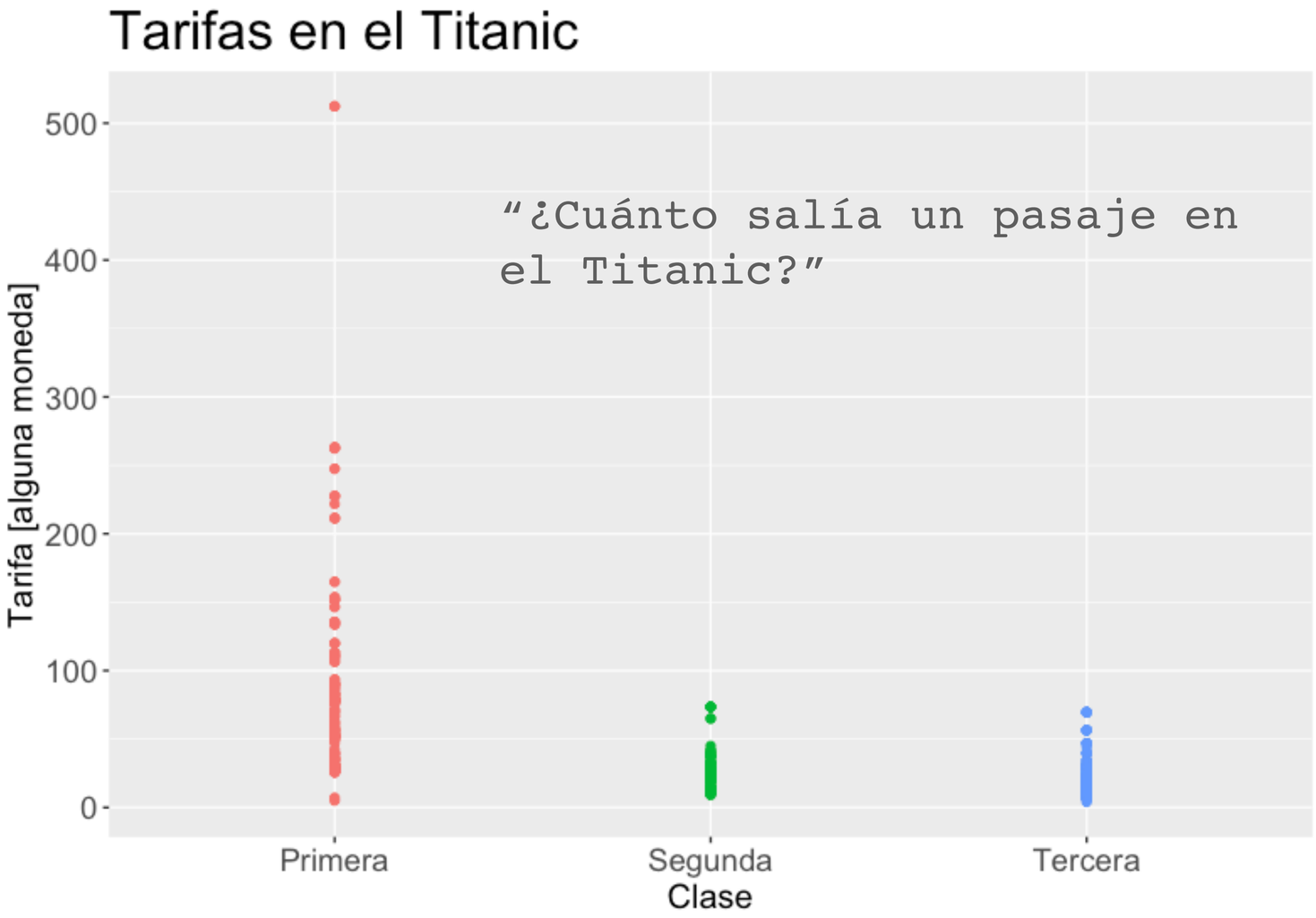
"¿Cuánto salía un pasaje en el Titanic?"





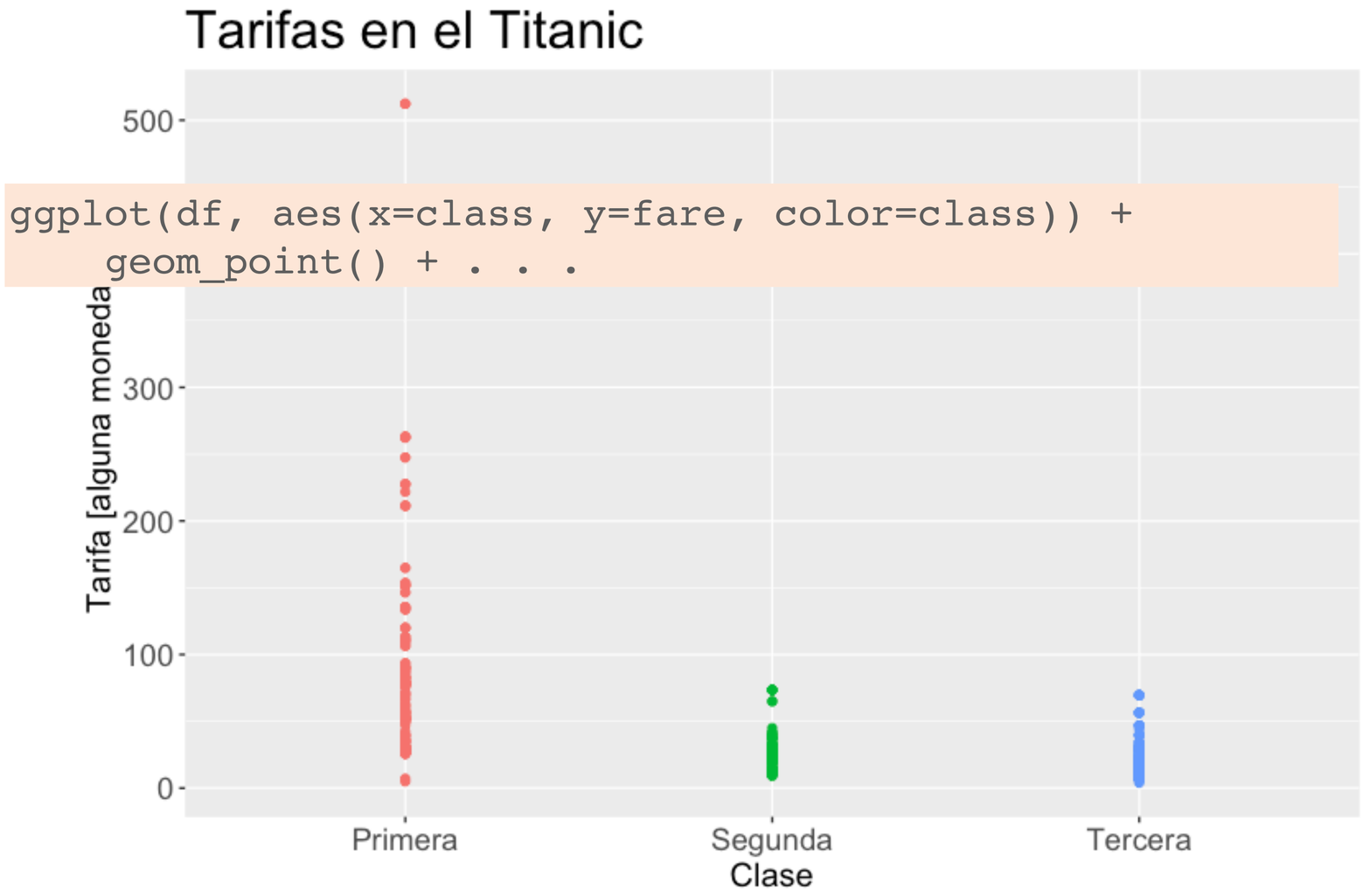
"¿Cuánto paga por el seguro médico una persona de 42 años no fumadora?"

Puntos usando una categórica como x (sí, ¿por qué no?)

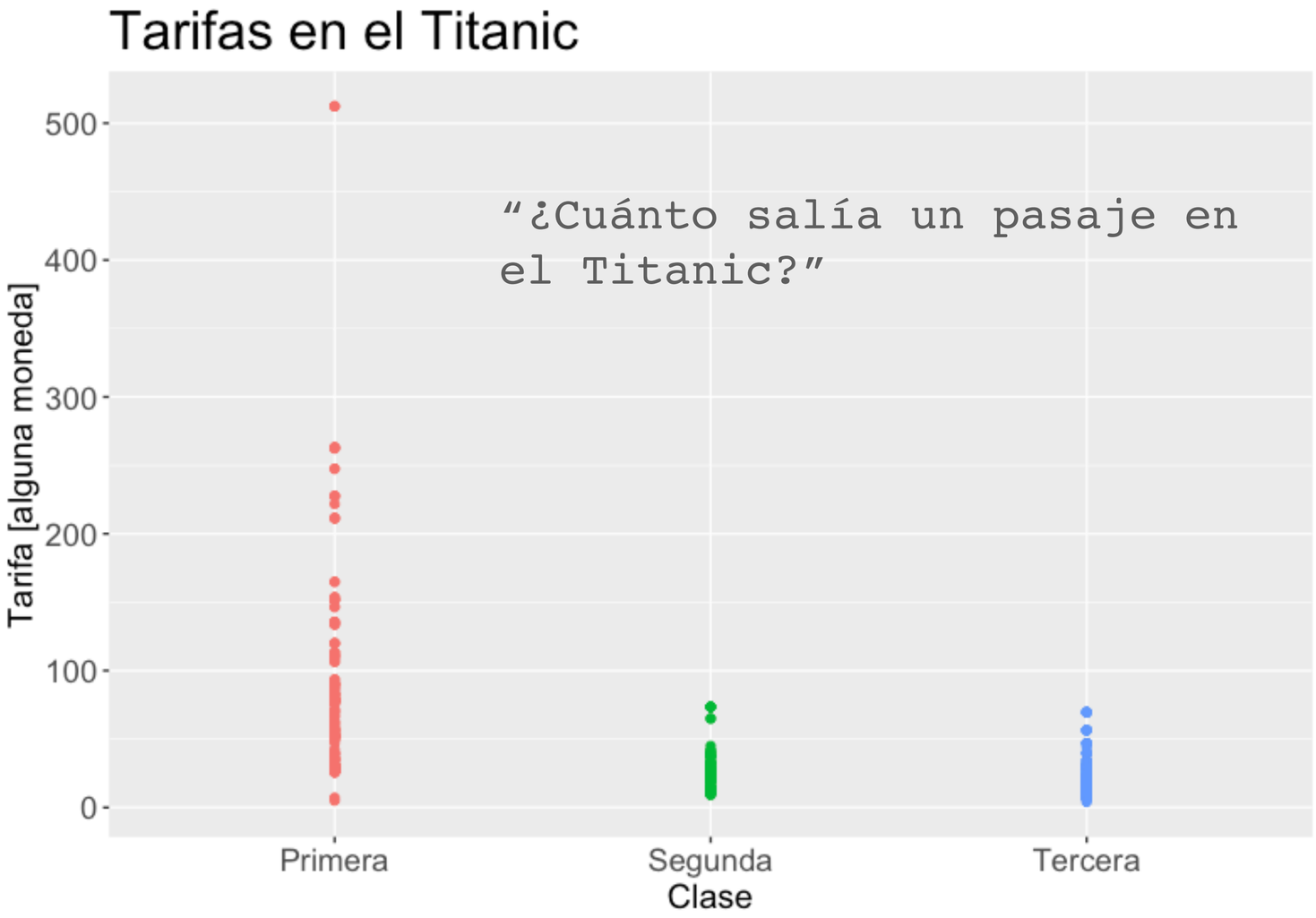


Puntos usando una categórica como x (sí, ¿por qué no?)

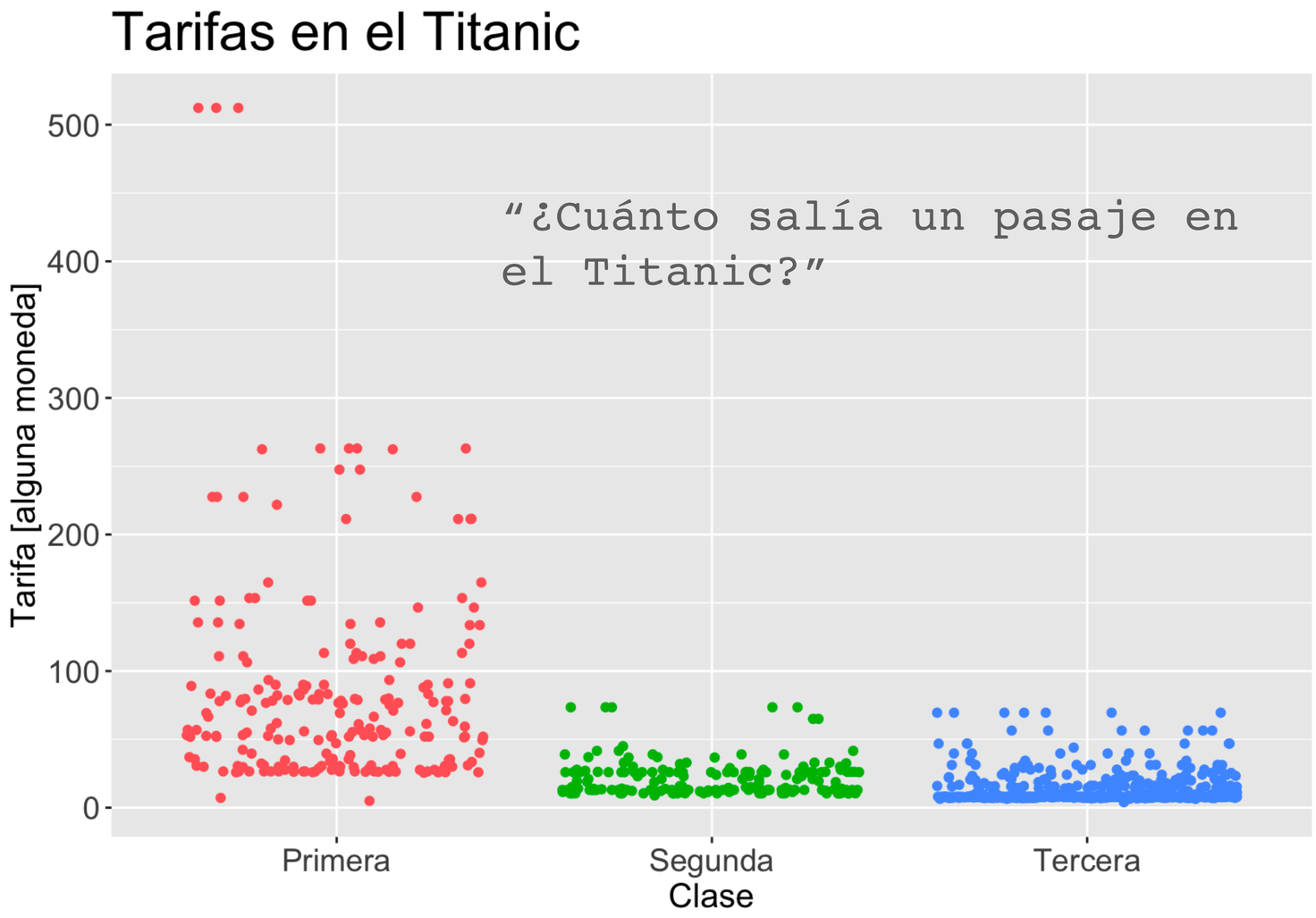

```
ggplot(df, aes(x=class, y=fare, color=class)) +
    geom_point() + . . .
```



Puntos usando una categórica como x (sí, ¿por qué no?)

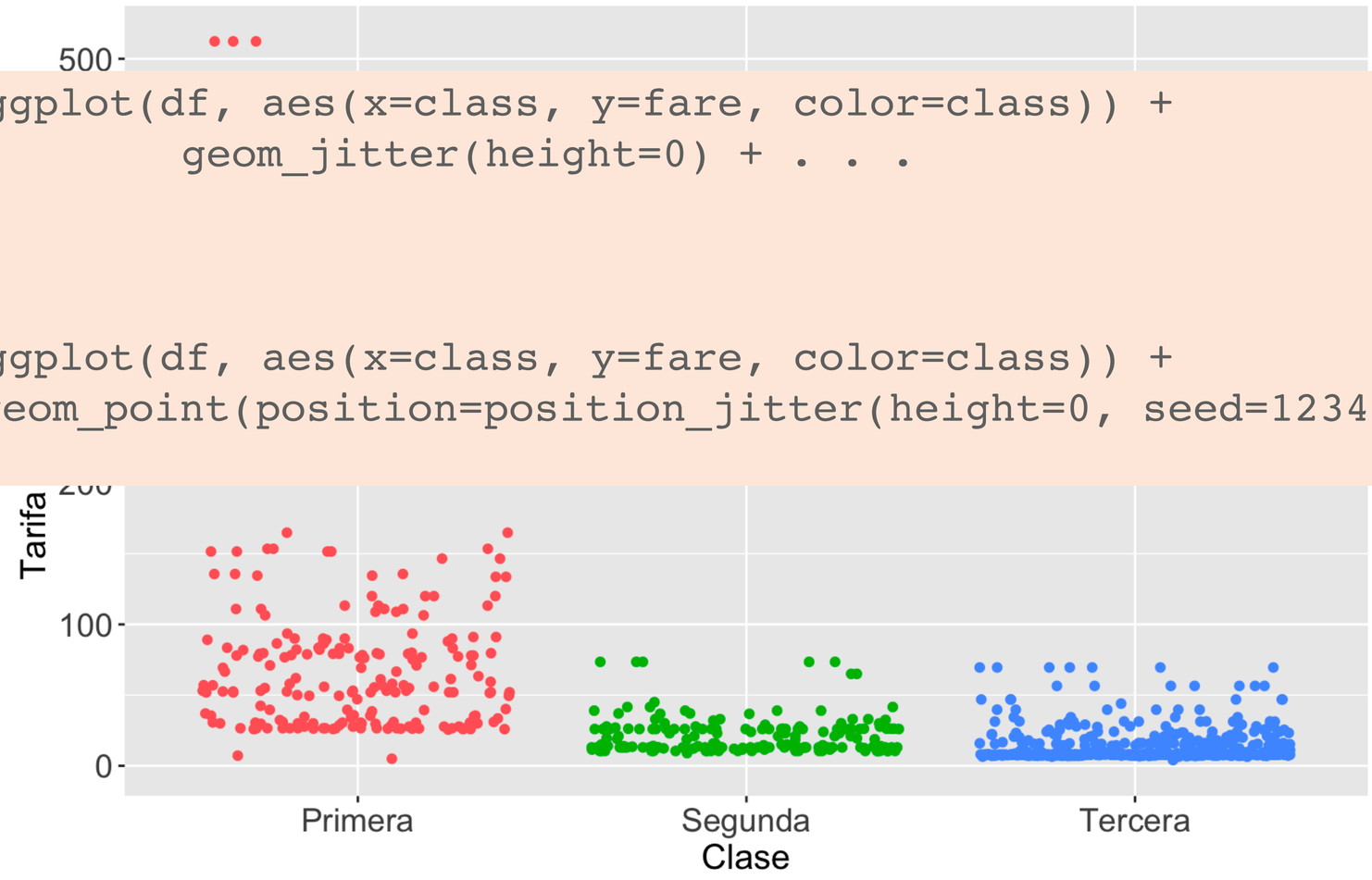


geom_jitter

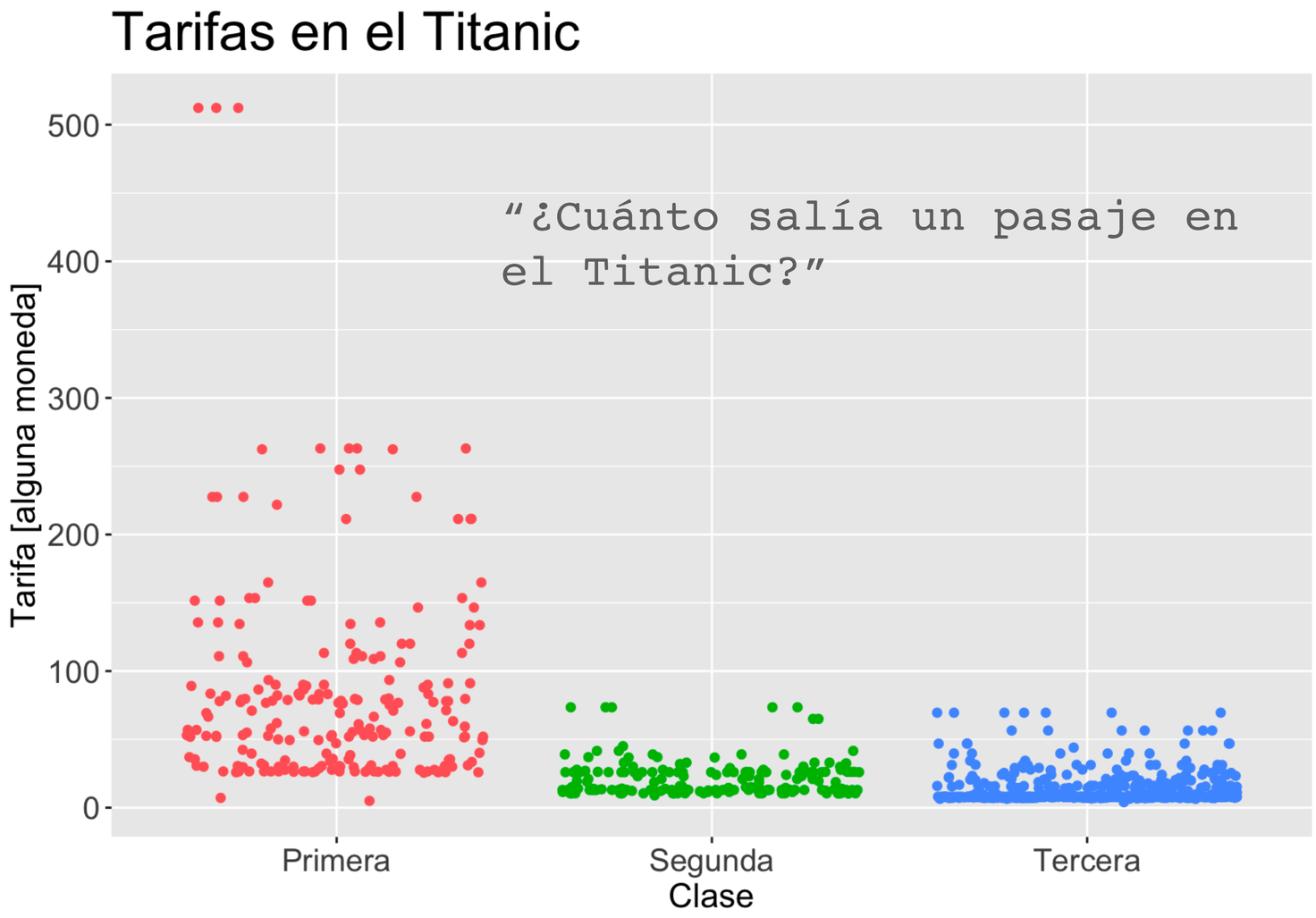

Tarifas en el Titanic
"¿Cuánto salía un pasaje en el Titanic?"
Tarifa [alguna moneda]
500
400
300
200
100
0
Primera
Segunda
Tercera
Clase

geom_jitter

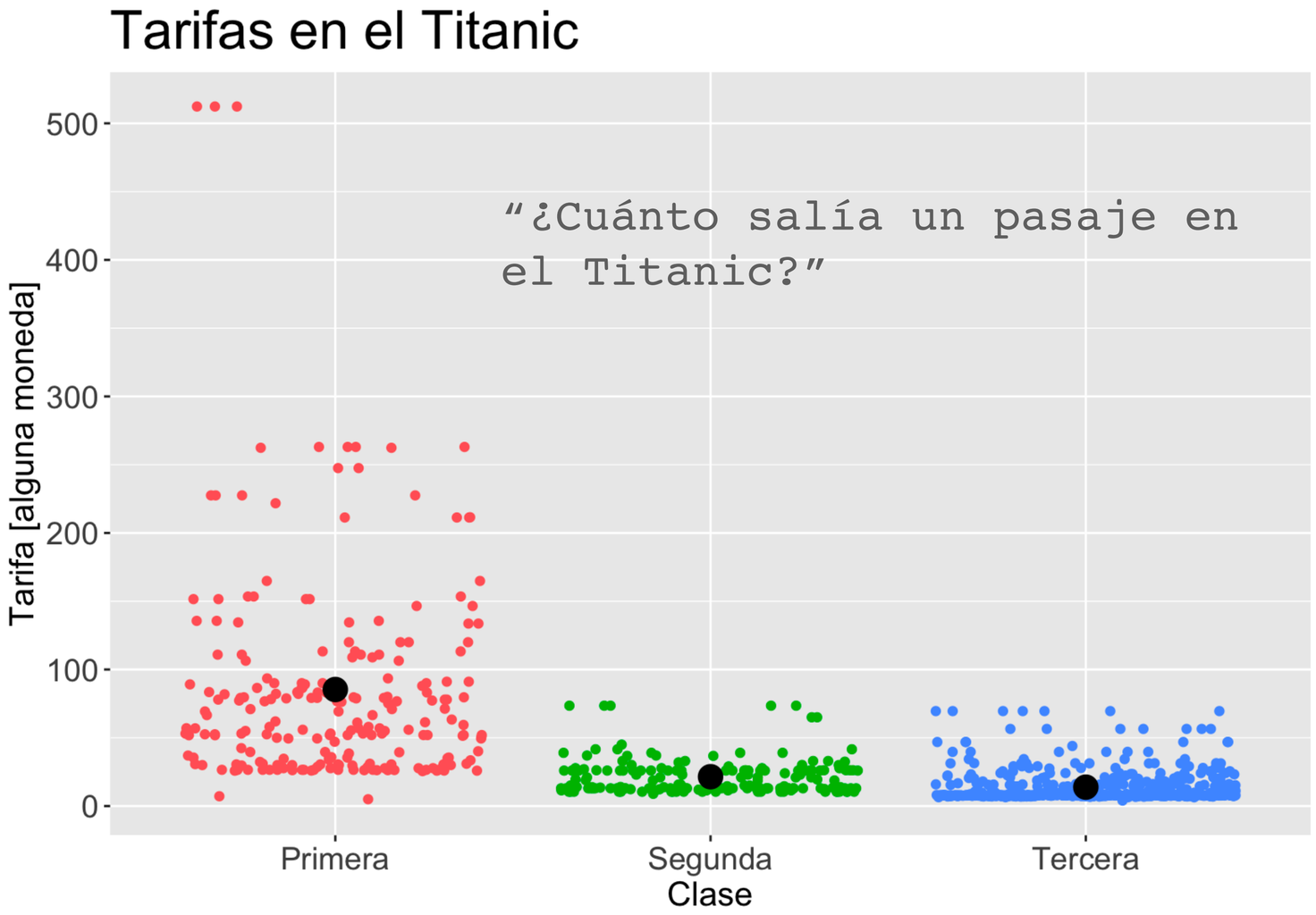

```
P <- ggplot(df, aes(x=class, y=fare, color=class)) +
             geom_jitter(height=0) + . . .

OR

p <- ggplot(df, aes(x=class, y=fare, color=class)) +
     geom_point(position=position_jitter(height=0, seed=1234)) + . . .
```



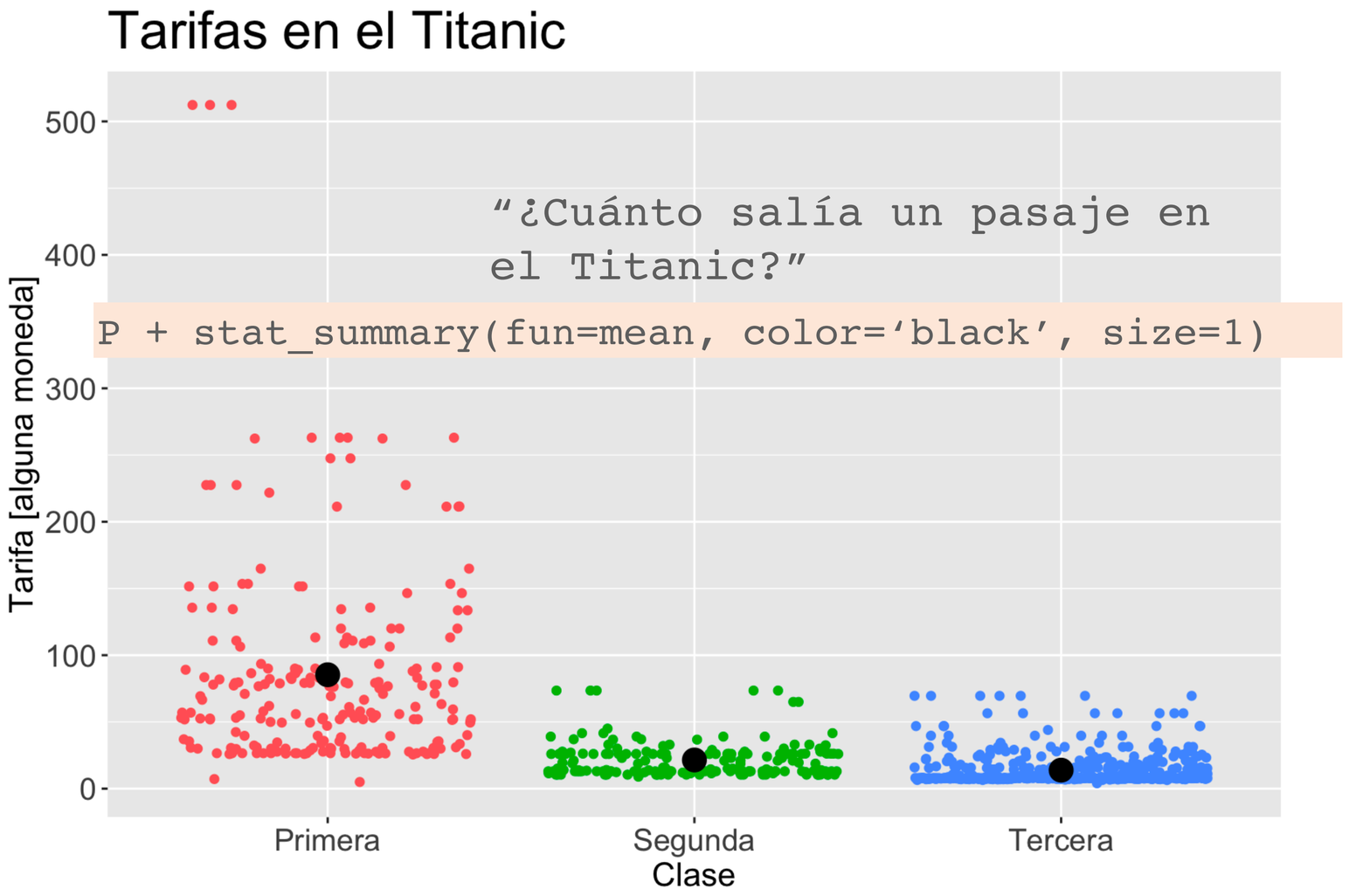
geom_jitter


Tarifas en el Titanic
"¿Cuánto salía un pasaje en el Titanic?"
Tarifa [alguna moneda]
500
400
300
200
100
0
Primera
Segunda
Tercera
Clase

Resumir los datos por clase



Resumir los datos por clase



Cómo se calcula


```
P + stat_summary(fun=mean, color='black', size=1)
```

	sex	age	siblings of the passenger	parents / children aboard the Titanic	fare	Port of Embarkation	class	who	alone
1	male	39.0	0	0	7.1800	S	First	man	TRUE
2	male	20.0	0	0	4.0125	C	Third	man	TRUE
3	male	33.0	0	0	5.0000	S	First	man	TRUE
4	male	61.0	0	0	6.2375	S	Third	man	TRUE
5	male	34.5	0	0	6.4375	C	Third	man	TRUE
6	male	43.0	0	0	6.4500	S	Third	man	TRUE
7	male	34.0	0	0	6.4958	S	Third	man	TRUE
8	male	18.0	1	0	6.4958	S	Third	man	FALSE
9	male	19.0	0	0	6.7500	Q	Third	man	TRUE
10	female	18.0	0	0	6.7500	Q	Third	woman	TRUE
11	male	*NA*	0	0	6.8583	Q	Third	man	TRUE

row names

Cómo se calcula


```
P + stat_summary(fun=mean, color='black', size=1)
```

	sex	age	siblings of the passenger	parents / children aboard the Titanic	fare	Port of Embarkation	class	who	alone
			row names						
1	male	39.0	0	0	7.1800	S	First	man	TRUE
2	male	20.0	0	0	4.0125	C	Third	man	TRUE
3	male	33.0	0	0	5.0000	S	First	man	TRUE
4	male	61.0	0	0	6.2375	S	Third	man	TRUE
5	male	34.5	0	0	6.4375	C	Third	man	TRUE
6	male	43.0	0	0	6.4500	S	Third	man	TRUE
7	male	34.0	0	0	6.4958	S	Third	man	TRUE
8	male	18.0	1	0	6.4958	S	Third	man	FALSE
9	male	19.0	0	0	6.7500	Q	Third	man	TRUE
10	female	18.0	0	0	6.7500	Q	Third	woman	TRUE
11	male	*NA*	0	0	6.8583	Q	Third	man	TRUE

	class	mean_fare
1	First	85.35230
2	Second	21.48816
3	Third	13.85217

Cómo se calcula


```
P + stat_summary(fun=mean, color='black', size=1)
```

	sex	age	siblings of the passenger	parents / children aboard the Titanic	fare	Port of Embarkation	class	who	alone
1	male	39.0	0	0	7.1800	S	First	man	TRUE
2	male	20.0	0	0	4.0125	C	Third	man	TRUE
3	male	33.0	0	0	5.0000	S	First	man	TRUE
4	male	61.0	0	0	6.2375	S	Third	man	TRUE
5	male	34.5	0	0	6.4375	C	Third	man	TRUE
6	male	43.0	0	0	6.4500	S	Third	man	TRUE
7	male	34.0	0	0	6.4958	S	Third	man	TRUE
8	male	18.0	1	0	6.4958	S	Third	man	FALSE
9	male	19.0	0	0	6.7500	Q	Third	man	TRUE
10	female	18.0	0	0	6.7500	Q	Third	woman	TRUE
11	male	*NA*	0	0	6.8583	Q	Third	man	TRUE

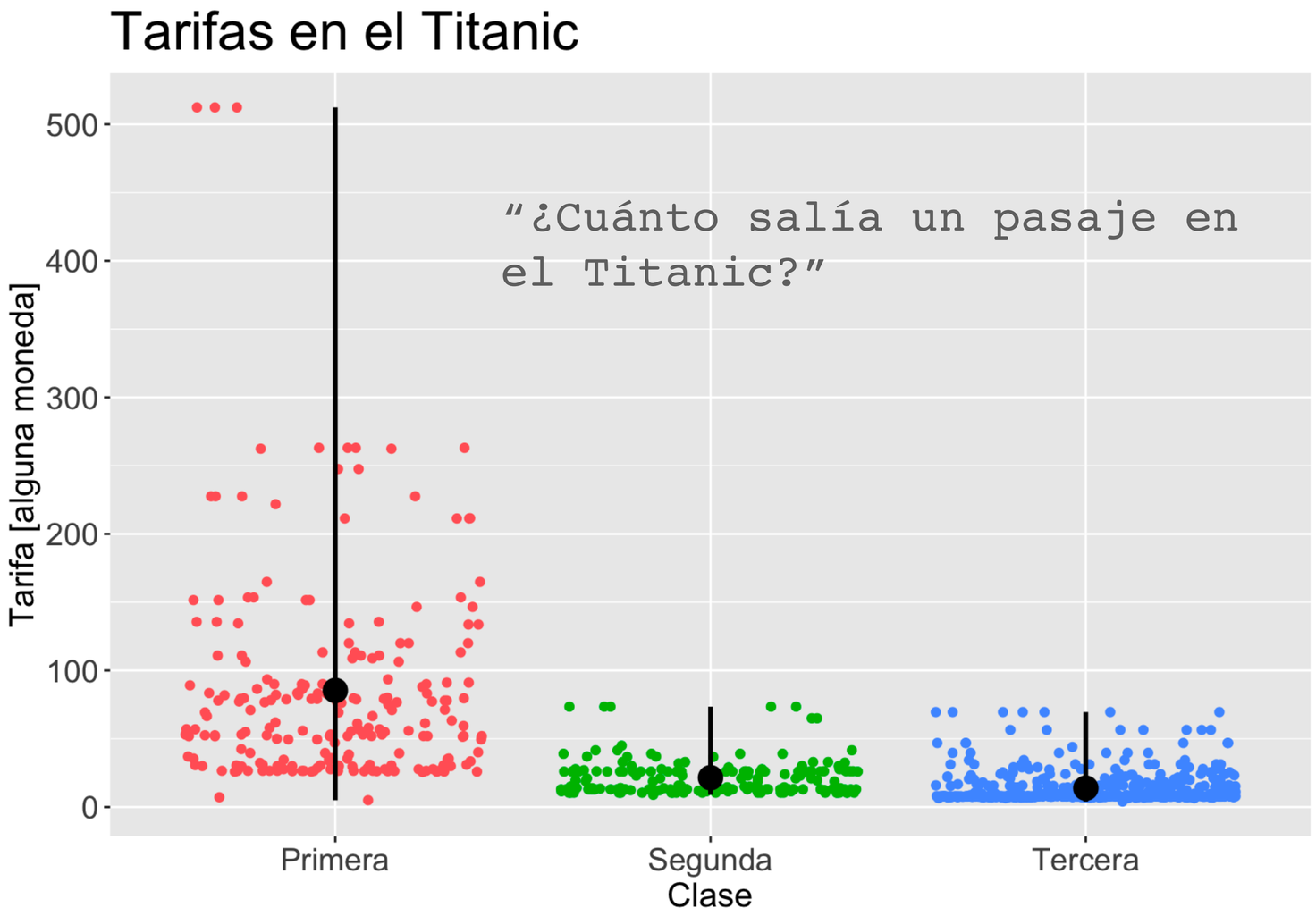
row names

```
df_summary <- df %>%
    group_by(class) %>%
    summarise(mean_fare = mean(fare, na.rm=T))
```

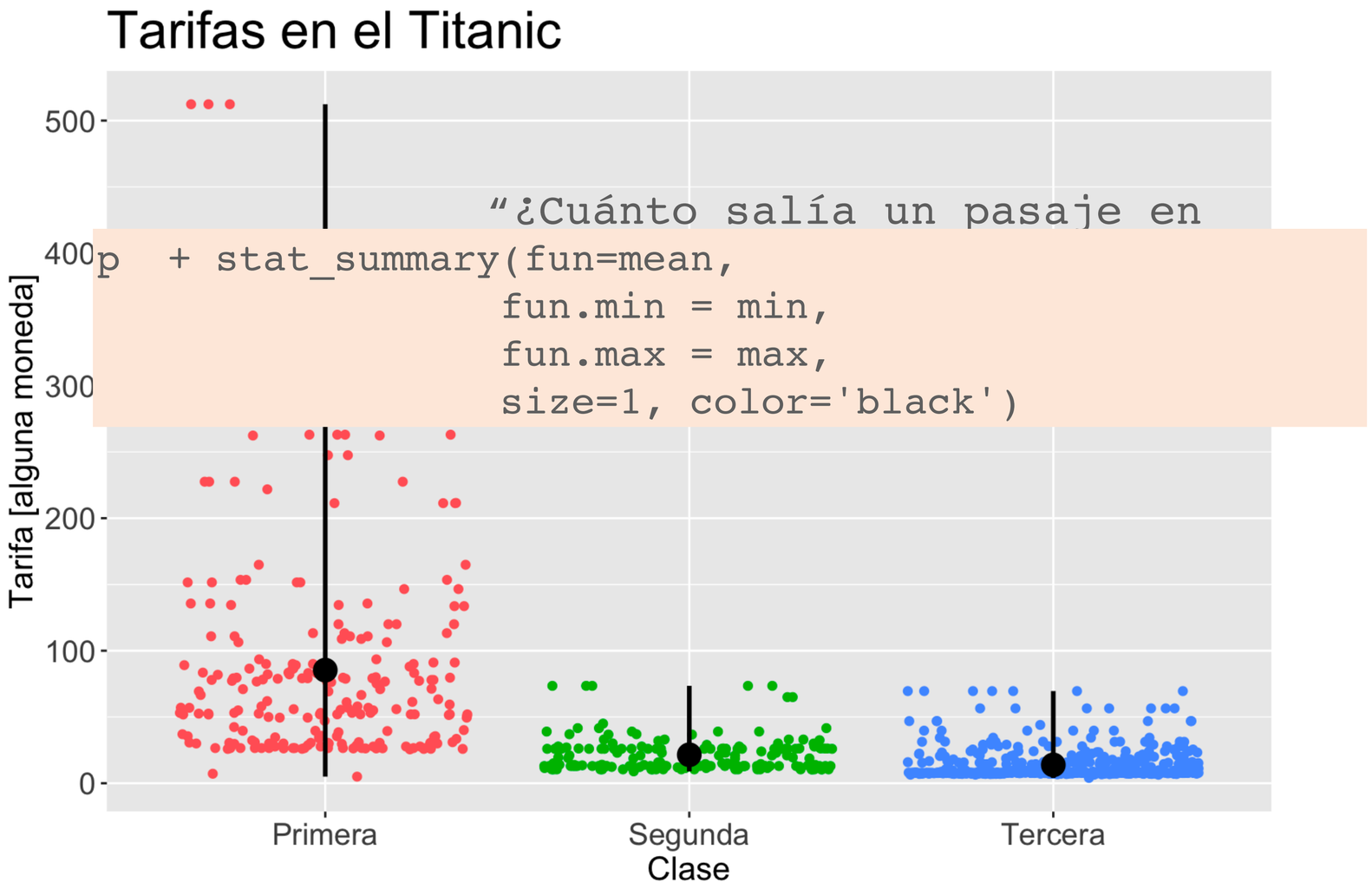
	class	mean_fare
1	First	85.35230
2	Second	21.48816
3	Third	13.85217

Alguna idea de la extensión

Licenciatura en Ciencia de Datos
ECYT_UNSAM



Alguna idea de la extensión

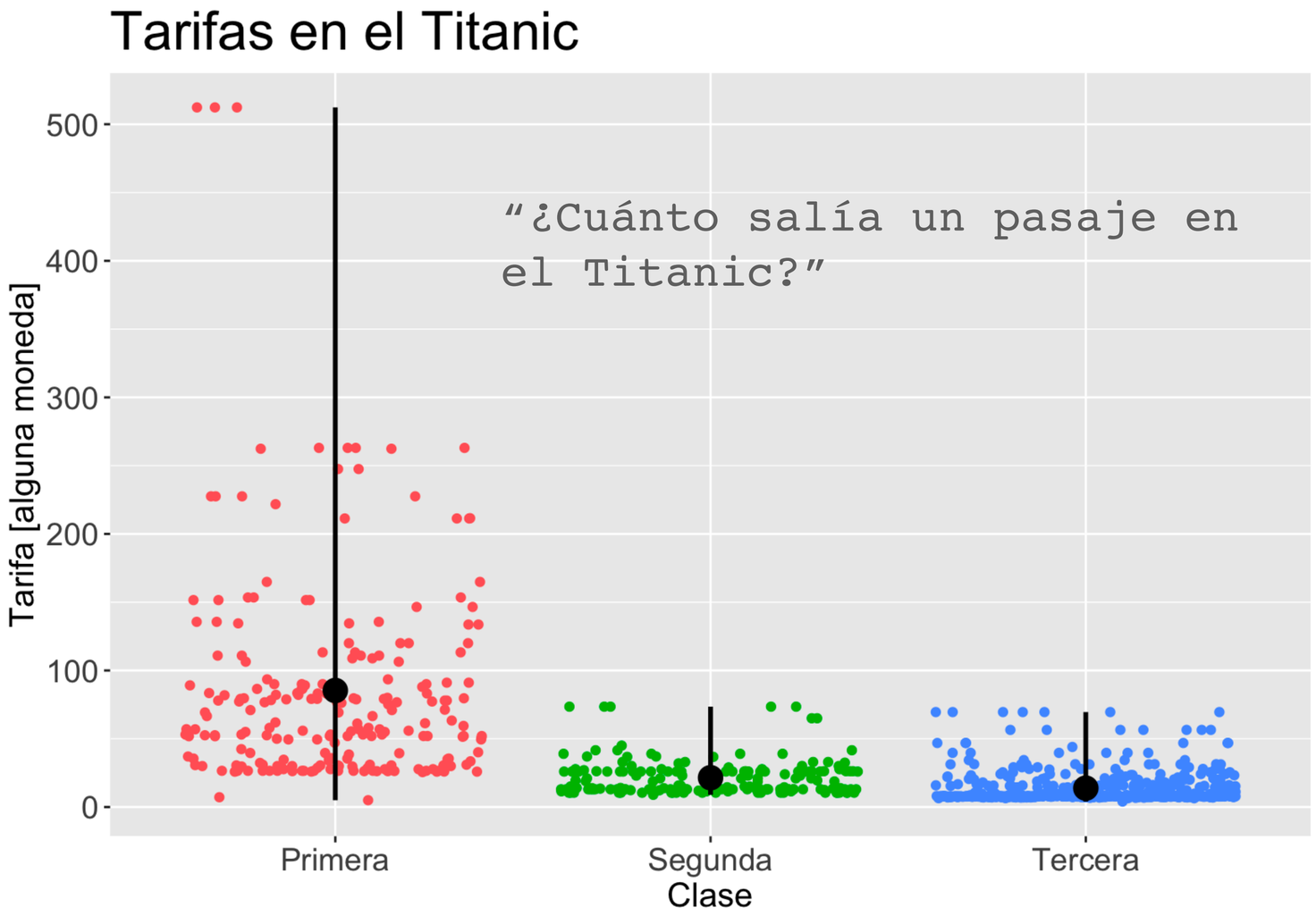



```
p  + stat_summary(fun=mean,
                  fun.min = min,
                  fun.max = max,
                  size=1, color='black')
```

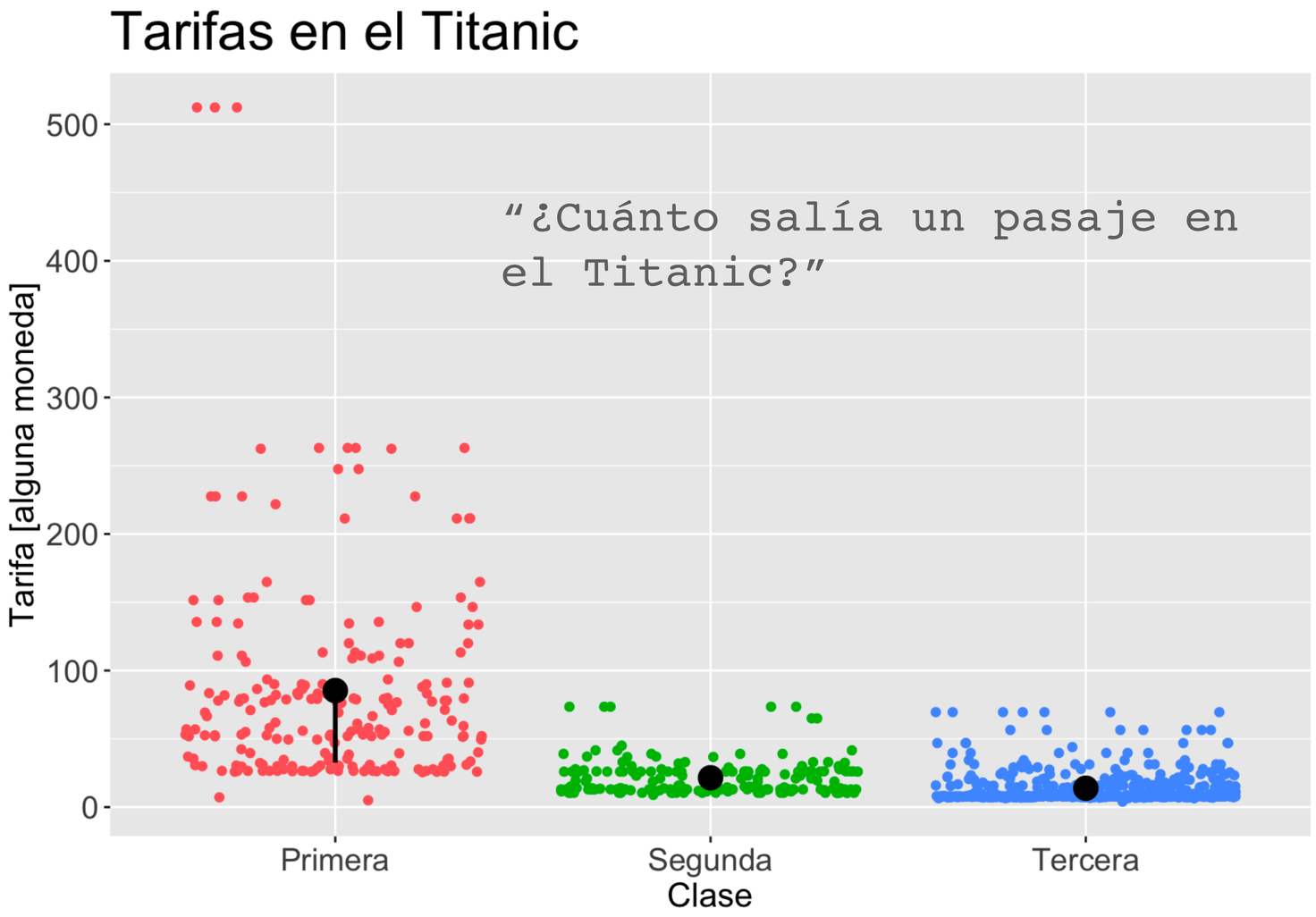
"¿Cuánto salía un pasaje en

Alguna idea de la extensión

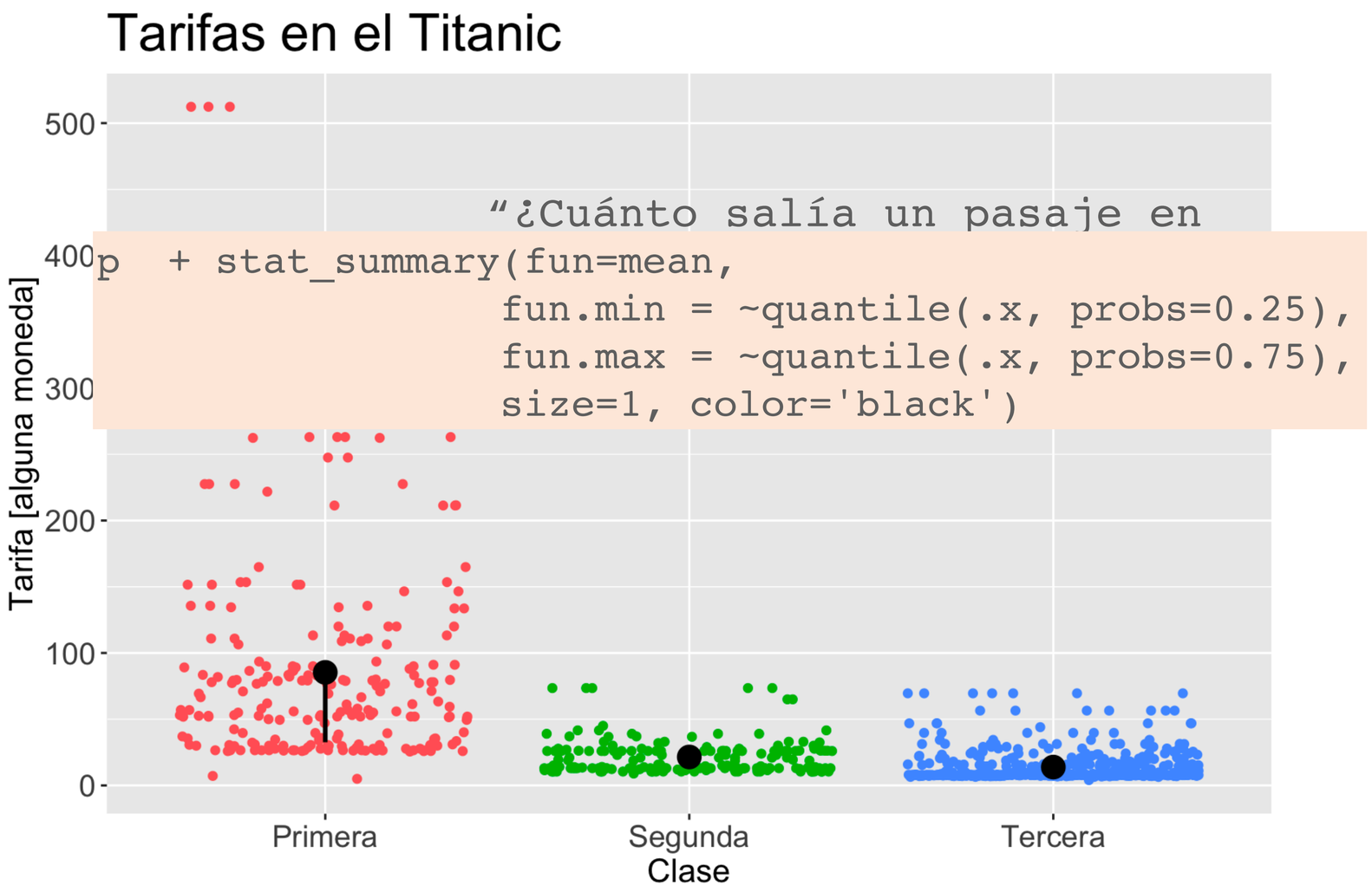
Licenciatura en Ciencia de Datos
ECYT_UNSAM



Cuantiles


Tarifas en el Titanic
"¿Cuánto salía un pasaje en el Titanic?"
Tarifa [alguna moneda]
500
400
300
200
100
0
Primera
Segunda
Tercera
Clase

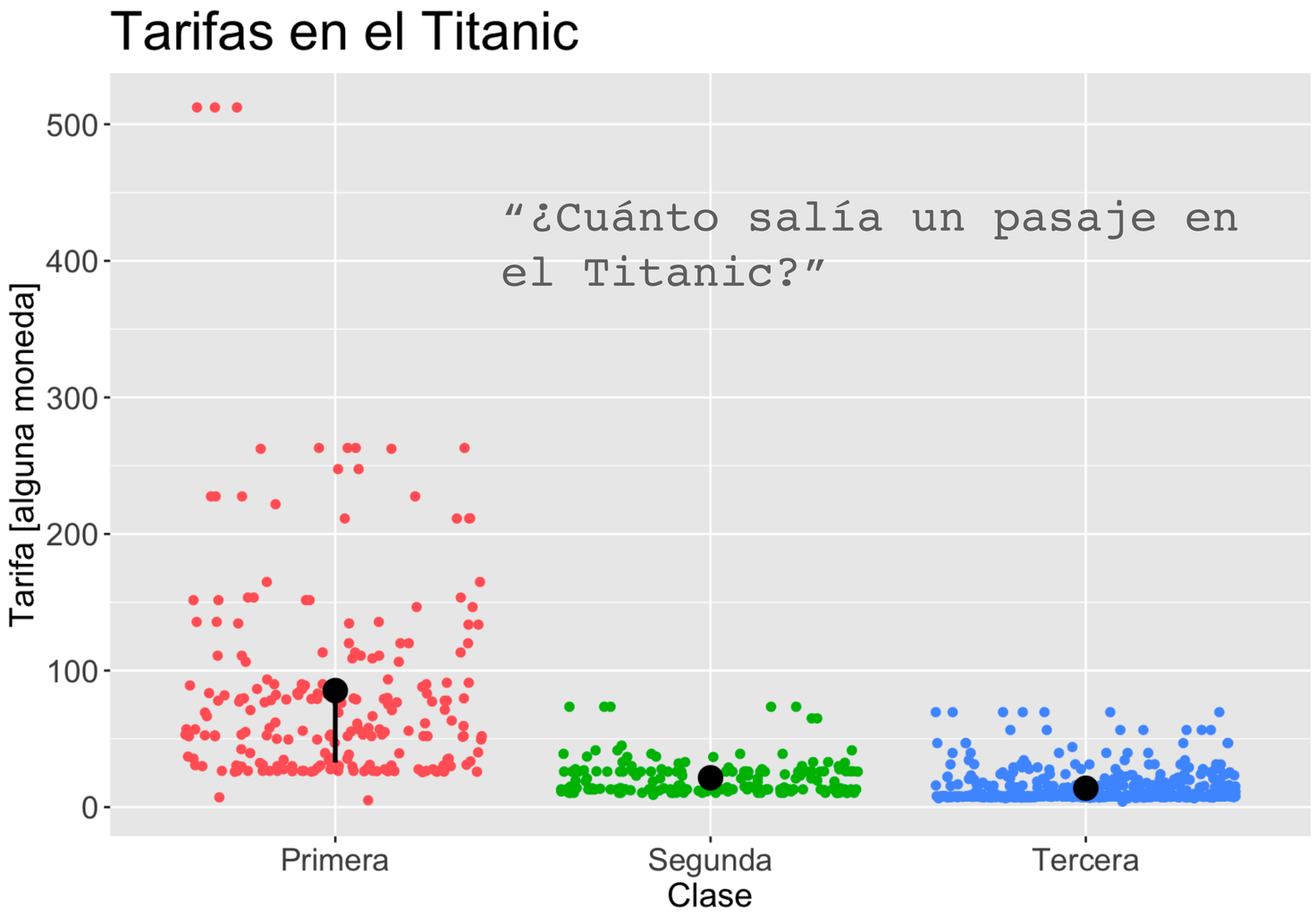
Cuantiles



“¿Cuánto salía un pasaje en

```
p  + stat_summary(fun=mean,
                  fun.min = ~quantile(.x, probs=0.25),
                  fun.max = ~quantile(.x, probs=0.75),
                  size=1, color='black')
```

Cuantiles


Tarifas en el Titanic
"¿Cuánto salía un pasaje en el Titanic?"
Tarifa [alguna moneda]
500
400
300
200
100
0
Primera
Segunda
Tercera
Clase

Cómo se calcula


```
P + stat_summary(fun=mean, color='black', size=1)
```

	sex	age	siblings of the passenger	parents / children aboard the Titanic	fare	Port of Embarkation	class	who	alone
1	male	39.0	0	0	7.1800	S	First	man	TRUE
2	male	20.0	0	0	4.0125	C	Third	man	TRUE
3	male	33.0	0	0	5.0000	S	First	man	TRUE
4	male	61.0	0	0	6.2375	S	Third	man	TRUE
5	male	34.5	0	0	6.4375	C	Third	man	TRUE
6	male	43.0	0	0	6.4500	S	Third	man	TRUE
7	male	34.0	0	0	6.4958	S	Third	man	TRUE
8	male	18.0	1	0	6.4958	S	Third	man	FALSE
9	male	19.0	0	0	6.7500	Q	Third	man	TRUE
10	female	18.0	0	0	6.7500	Q	Third	woman	TRUE
11	male	*NA*	0	0	6.8583	Q	Third	man	TRUE

row names

Cómo se calcula


```
P + stat_summary(fun=mean, color='black', size=1)
```

	sex	age	siblings of the passenger	parents / children aboard the Titanic	fare	Port of Embarkation	class	who	alone
			row names						
1	male	39.0	0	0	7.1800	S	First	man	TRUE
2	male	20.0	0	0	4.0125	C	Third	man	TRUE
3	male	33.0	0	0	5.0000	S	First	man	TRUE
4	male	61.0	0	0	6.2375	S	Third	man	TRUE
5	male	34.5	0	0	6.4375	C	Third	man	TRUE
6	male	43.0	0	0	6.4500	S	Third	man	TRUE
7	male	34.0	0	0	6.4958	S	Third	man	TRUE
8	male	18.0	1	0	6.4958	S	Third	man	FALSE
9	male	19.0	0	0	6.7500	Q	Third	man	TRUE
10	female	18.0	0	0	6.7500	Q	Third	woman	TRUE
11	male	*NA*	0	0	6.8583	Q	Third	man	TRUE

class	mean_fare	Q25	Q75
First	85.35230	32.6156	93.5
Second	21.48816	13.0000	26.0
Third	13.85217	7.7500	15.5

Cómo se calcula


```
P + stat_summary(fun=mean, color='black', size=1)
```

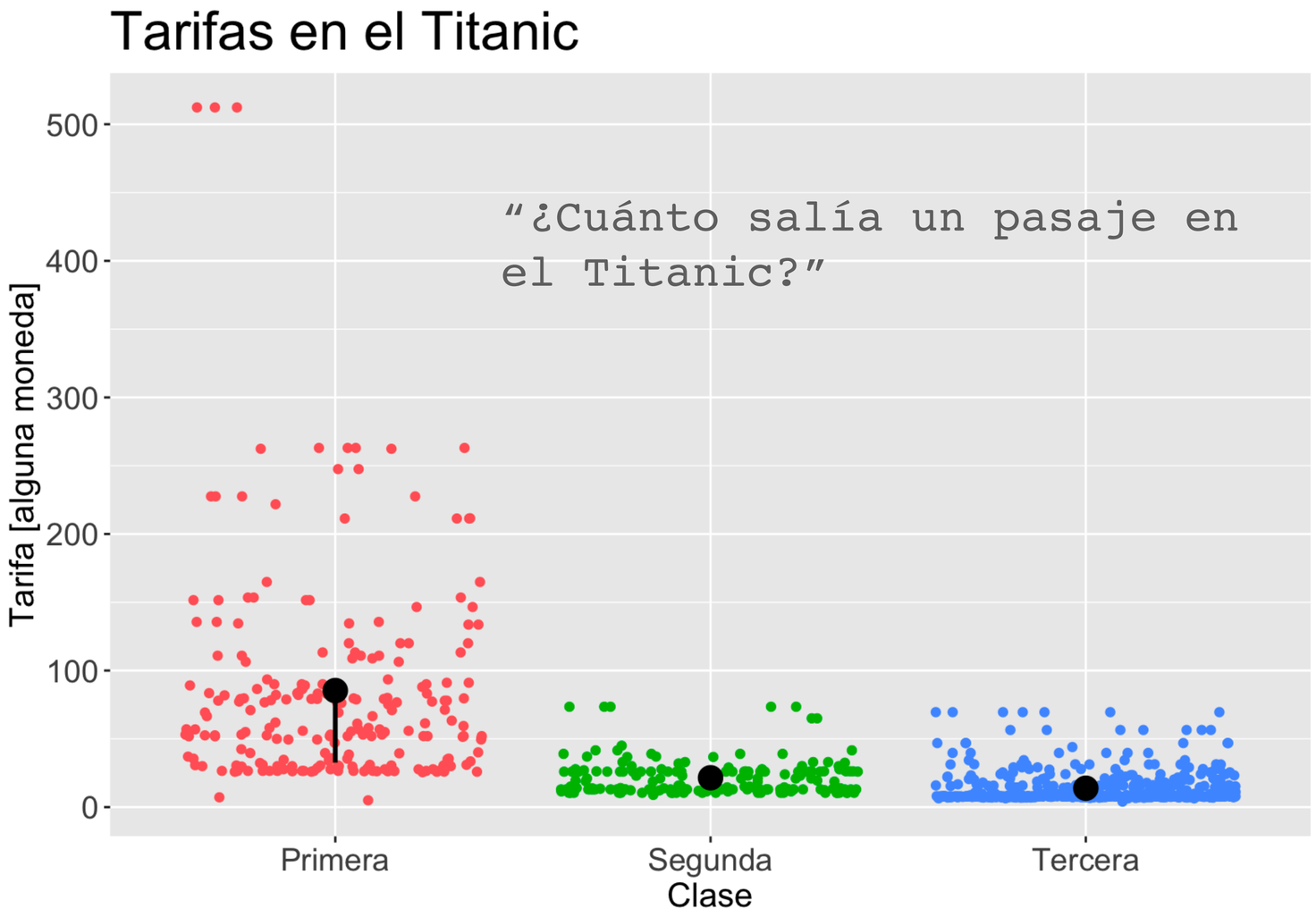
	sex	age	siblings of the passenger	parents / children aboard the Titanic	fare	Port of Embarkation	class	who	alone
1	male	39.0	0	0	7.1800	S	First	man	TRUE
2	male	20.0	0	0	4.0125	C	Third	man	TRUE
3	male	33.0	0	0	5.0000	S	First	man	TRUE
4	male	61.0	0	0	6.2375	S	Third	man	TRUE
5	male	34.5	0	0	6.4375	C	Third	man	TRUE
6	male	43.0	0	0	6.4500	S	Third	man	TRUE
7	male	34.0	0	0	6.4958	S	Third	man	TRUE
8	male	18.0	1	0	6.4958	S	Third	man	FALSE
9	male	19.0	0	0	6.7500	Q	Third	man	TRUE
10	female	18.0	0	0	6.7500	Q	Third	woman	TRUE
11	male	*NA*	0	0	6.8583	Q	Third	man	TRUE

row names

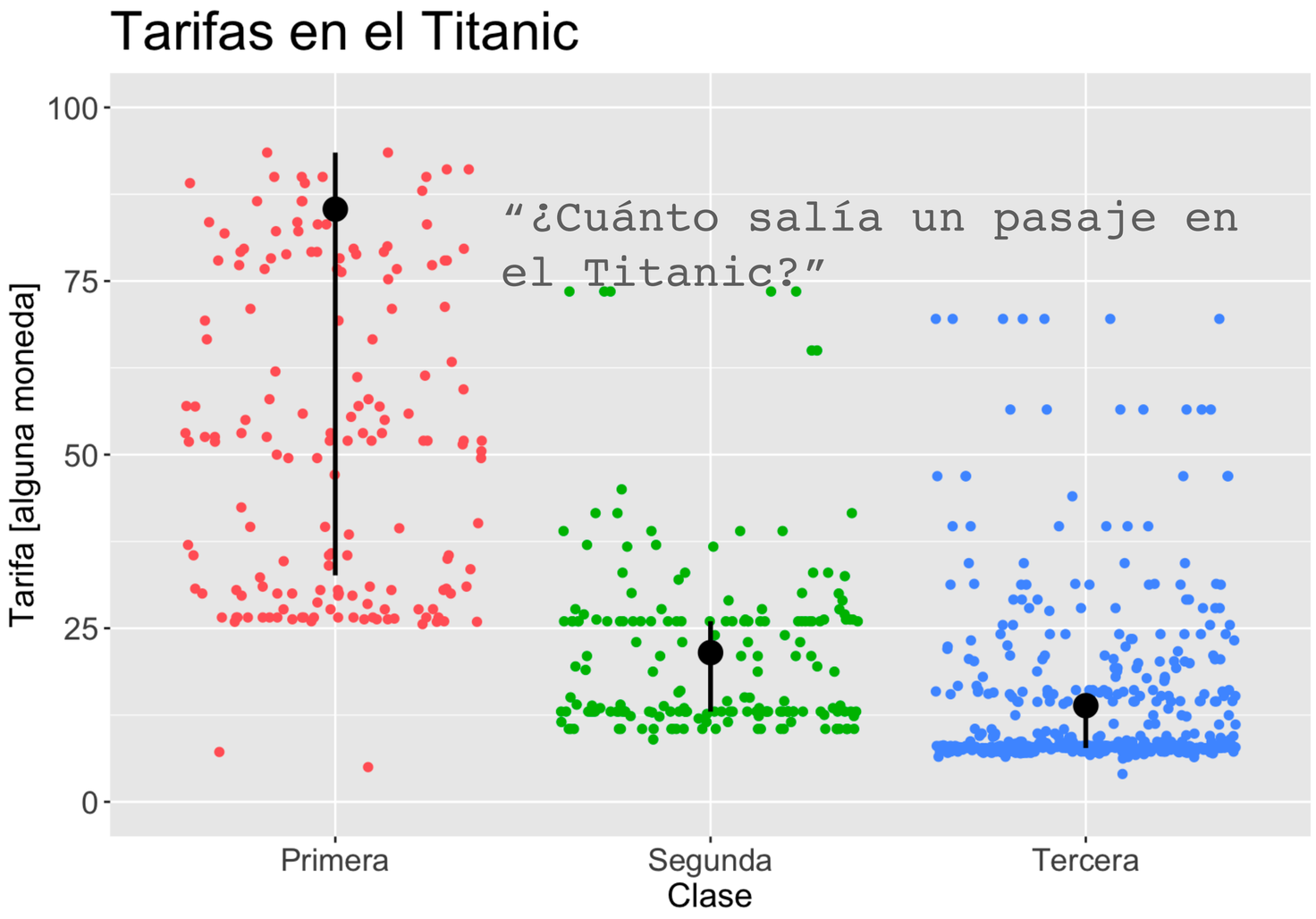
```
df_summary <- df %>%
    group_by(class) %>%
    summarise(mean_fare = mean(fare, na.rm=T),
                        Q25 = quantile(fare, probs=0.25, na.rm=T),
                        Q75 = quantile(fare, probs=0.75, na.rm=T))
```

class	mean_fare	Q25	Q75
First	85.35230	32.6156	93.5
Second	21.48816	13.0000	26.0
Third	13.85217	7.7500	15.5

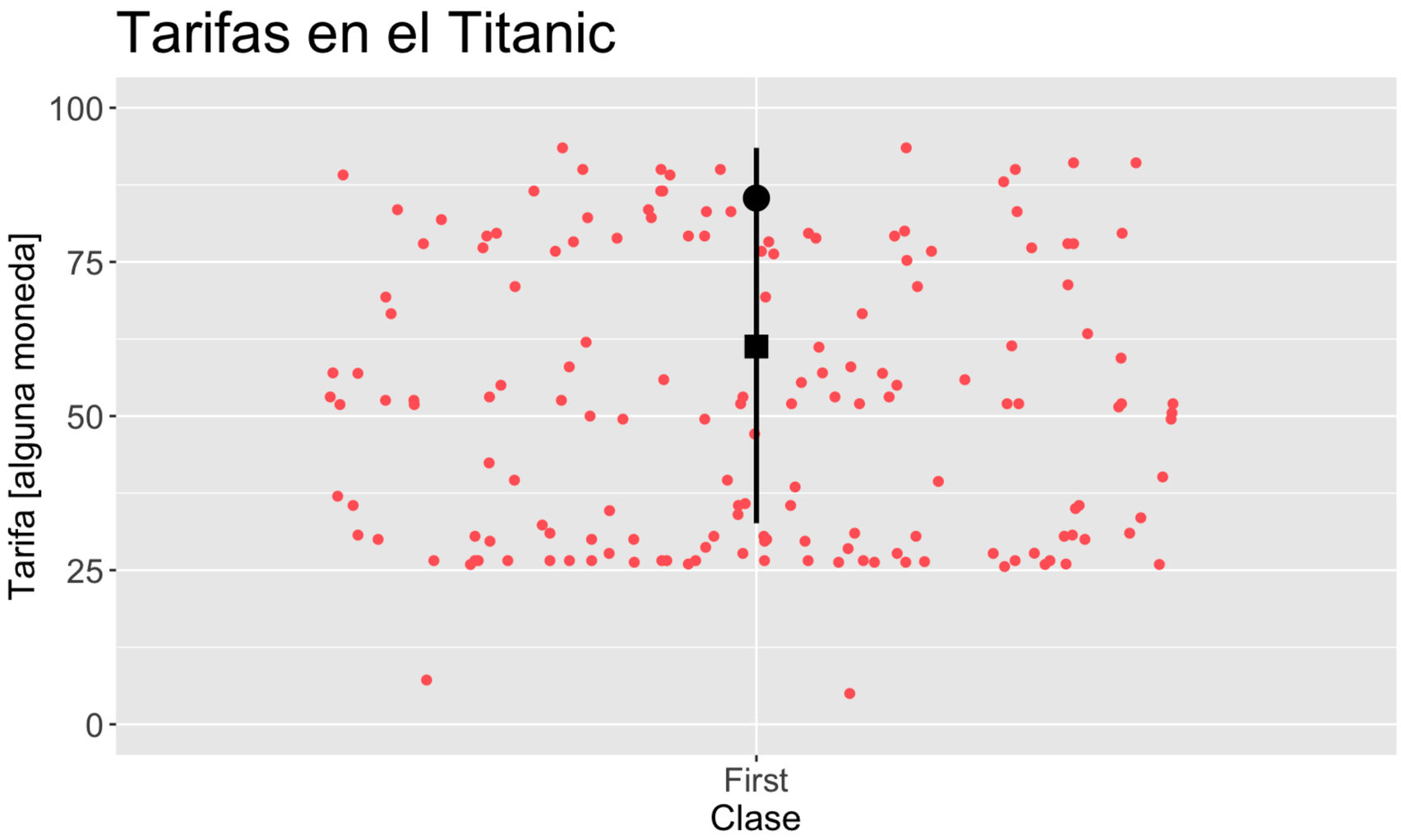
Cuantiles


Tarifas en el Titanic
"¿Cuánto salía un pasaje en el Titanic?"
Tarifa [alguna moneda]
500
400
300
200
100
0
Primera
Segunda
Tercera
Clase

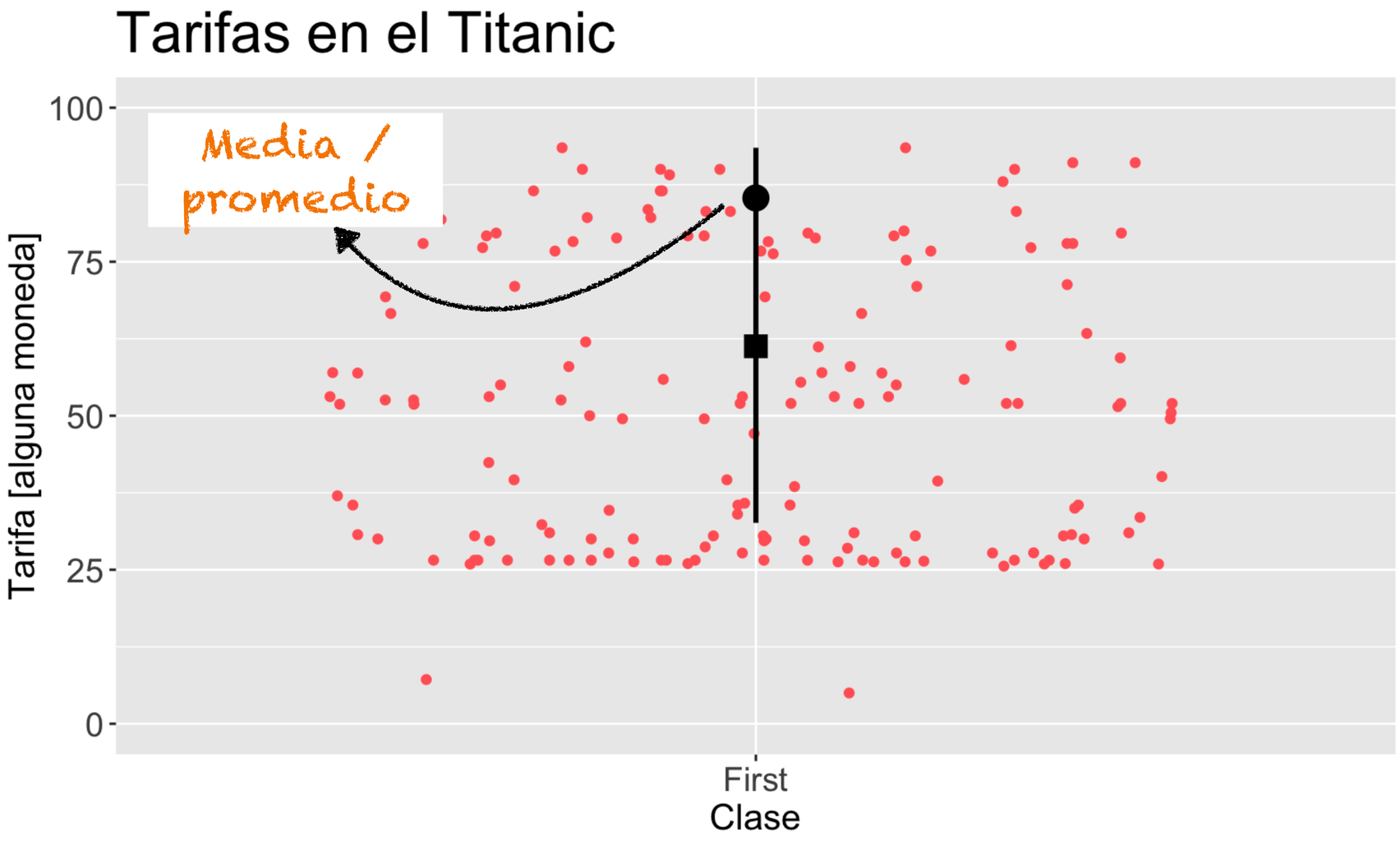
Cuantiles


Tarifas en el Titanic
"¿Cuánto salía un pasaje en el Titanic?"
Tarifa [alguna moneda]
100
75
50
25
0
Primera
Segunda
Tercera
Clase

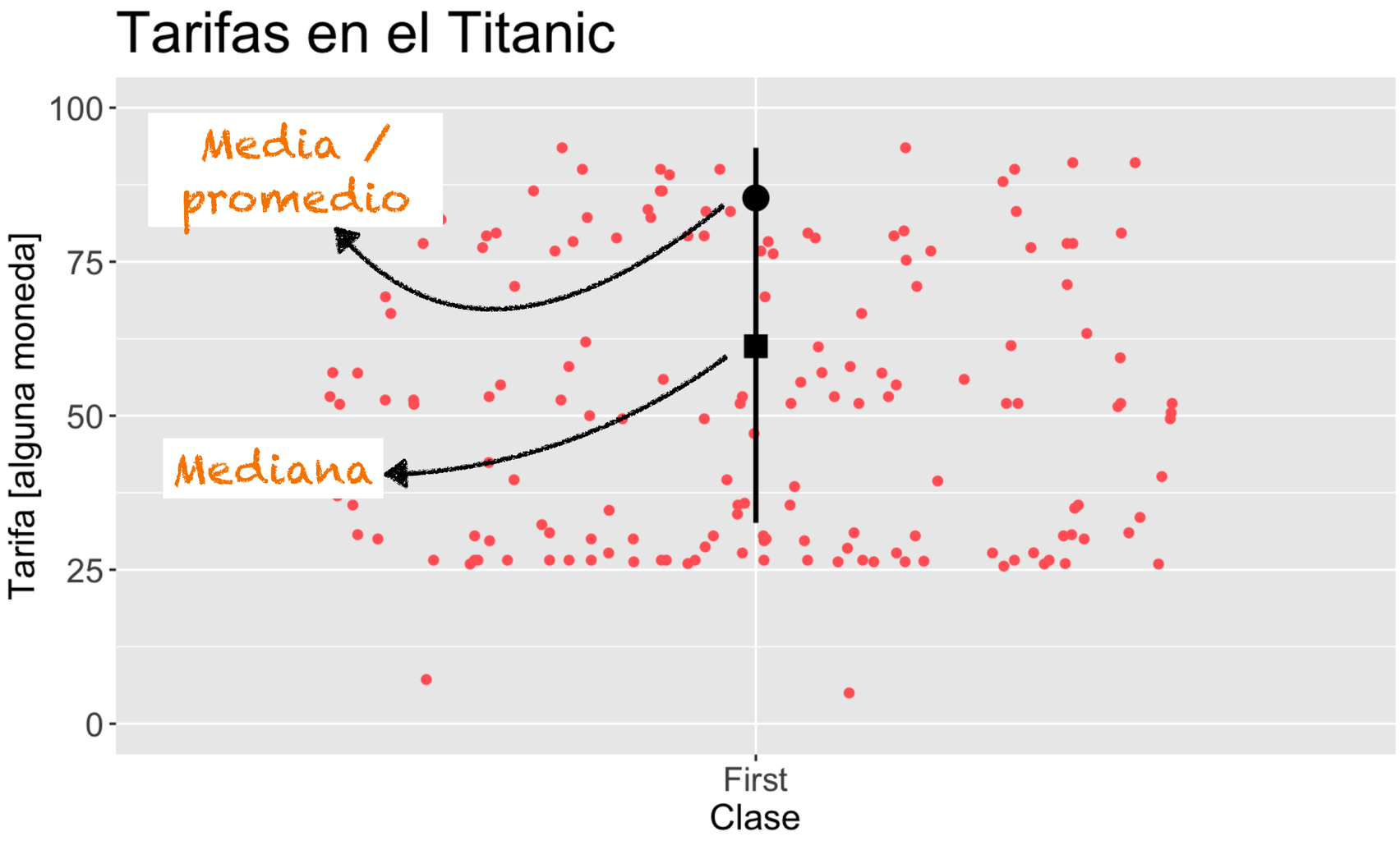
Más métricas



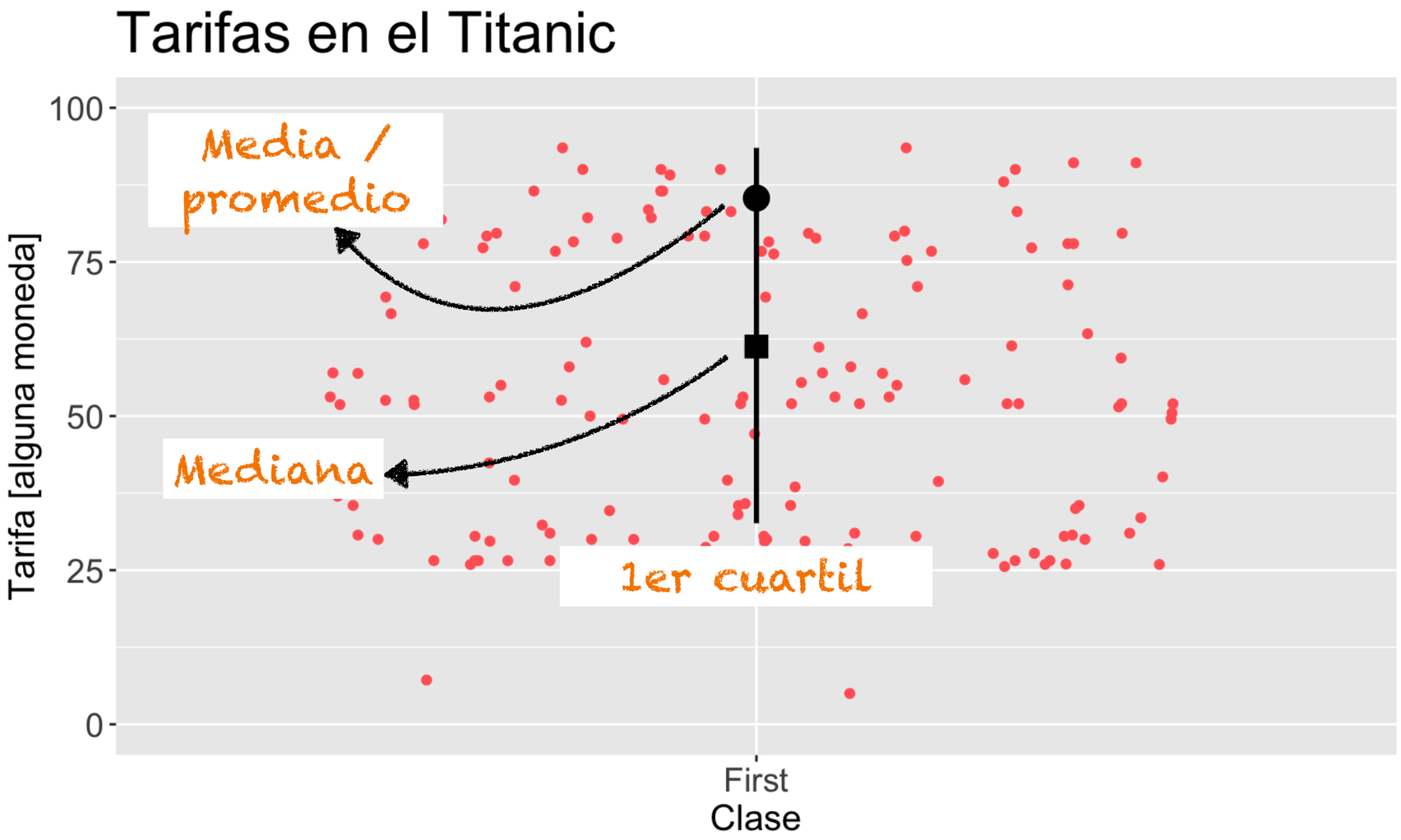
Más métricas



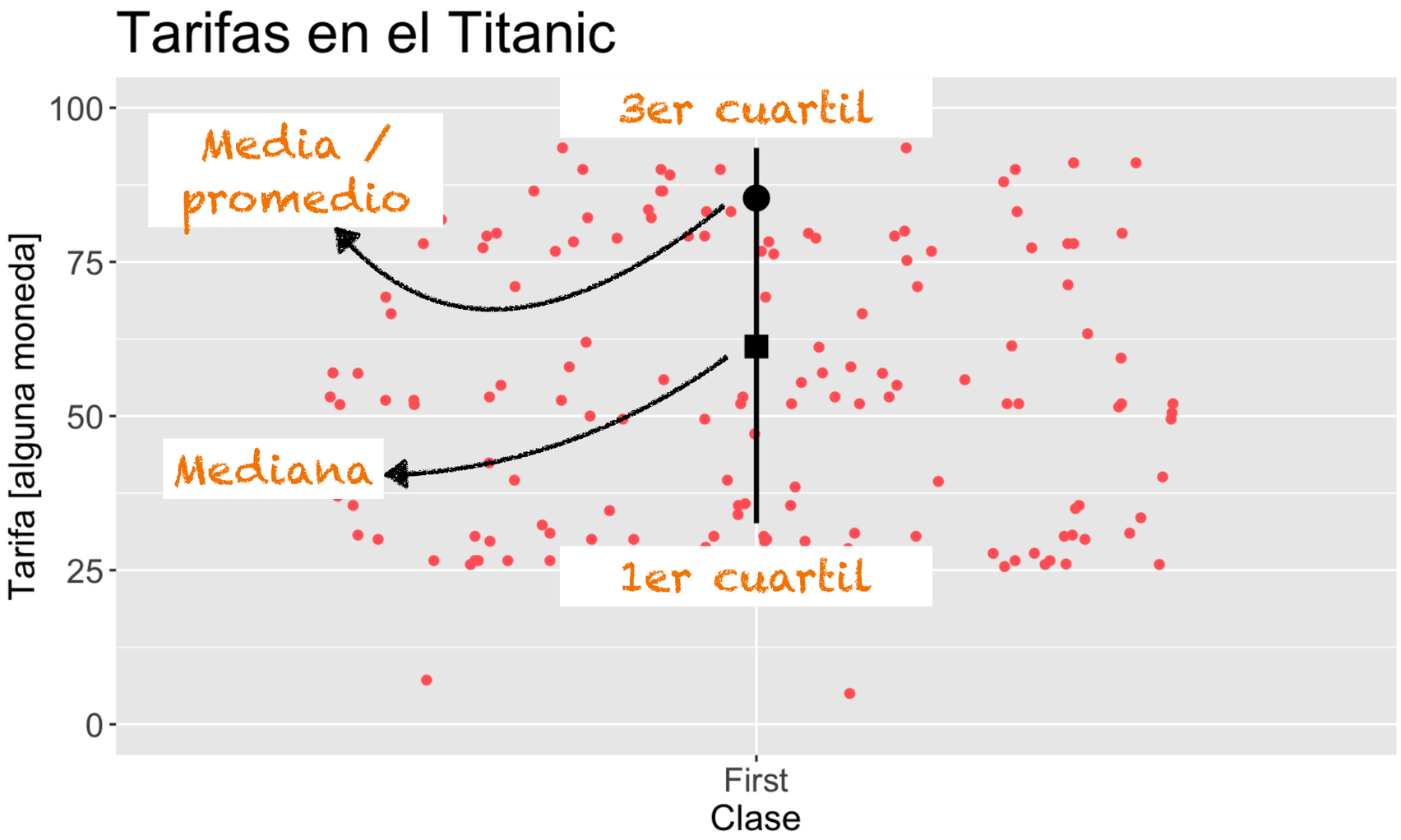
Más métricas



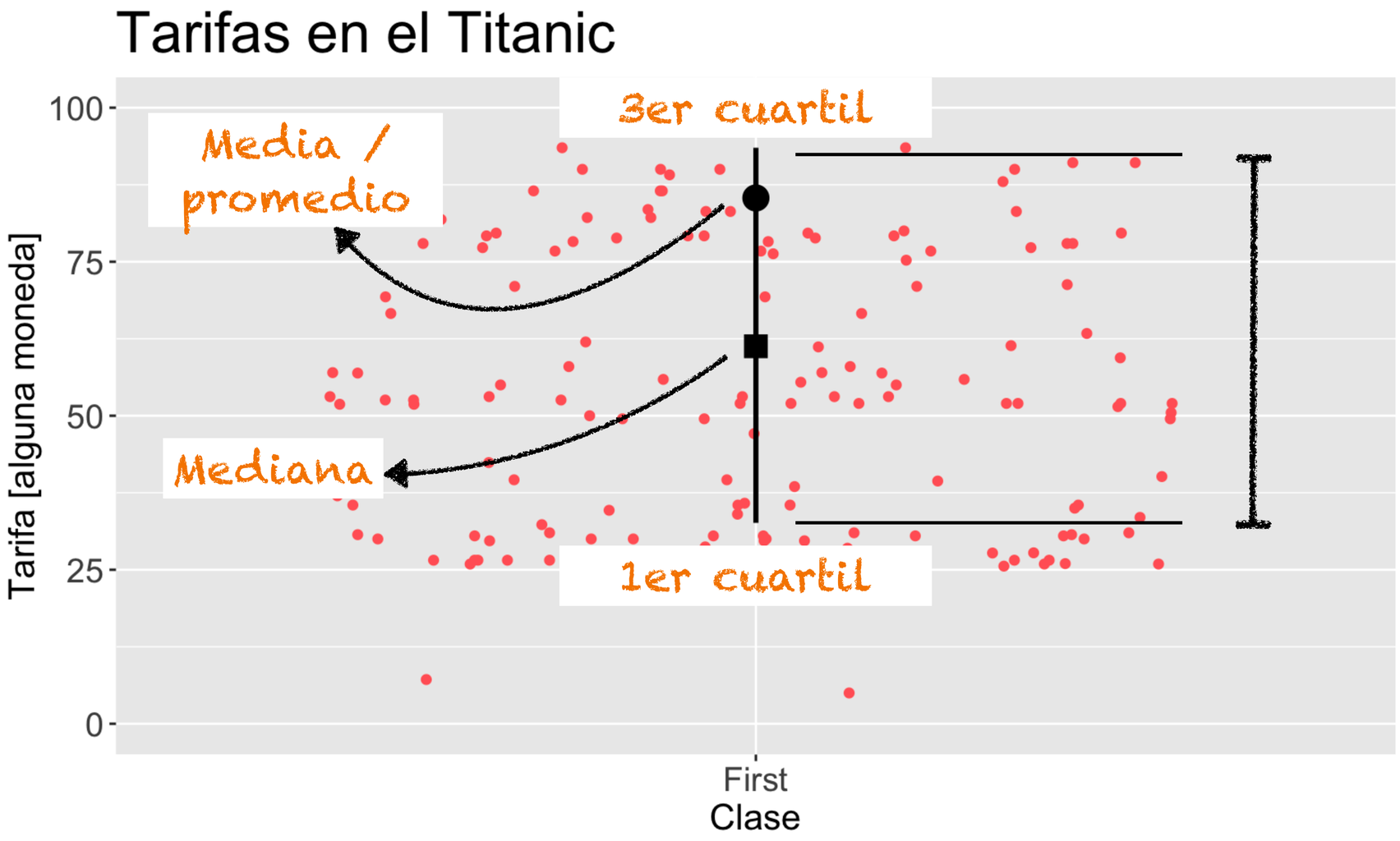
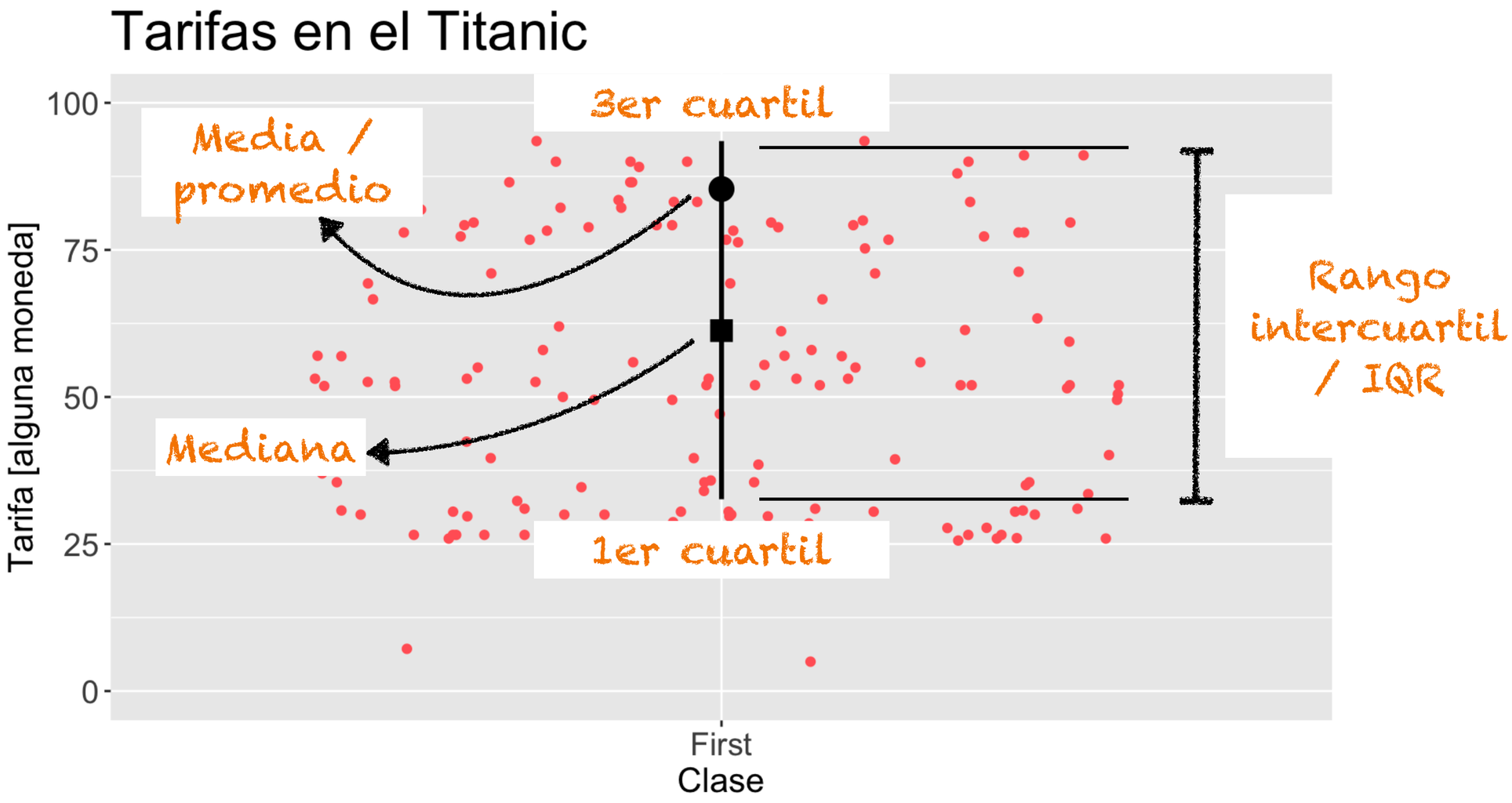
Más métricas



Más métricas



Más métricas



Más métricas