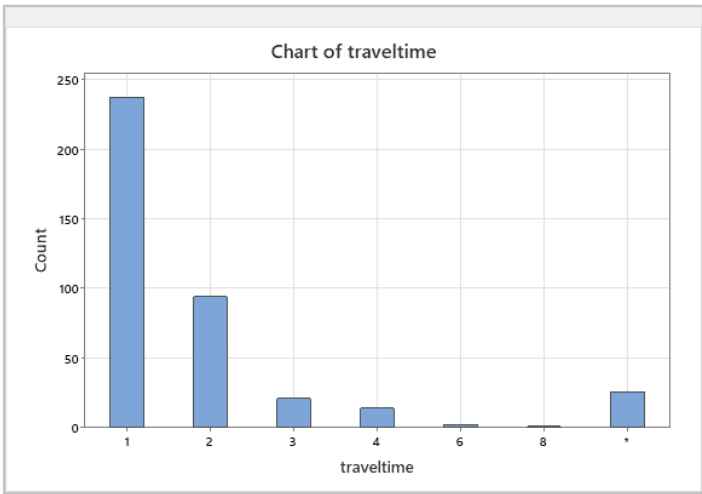


Manejo de Datos

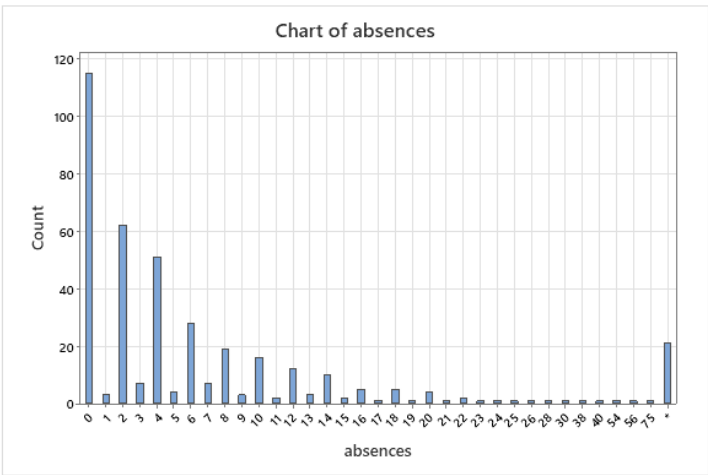
1.Identificar el porcentaje de datos faltantes.



Tally

traveltime	Count	Percent
1	237	64.23
2	94	25.47
3	21	5.69
4	14	3.79
6	2	0.54
8	1	0.27
N=	369	
*=	26	

Traveltime:  $26/395 = 6.58\%$

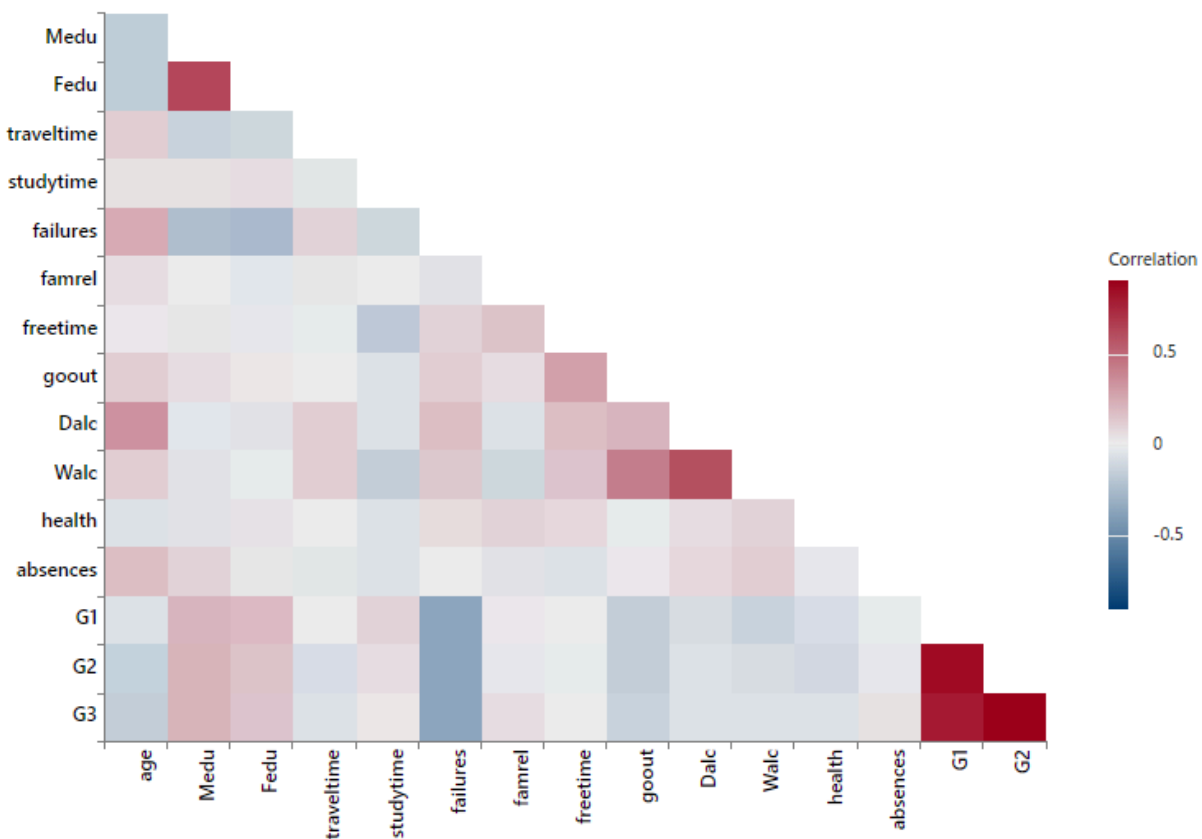


Tally

absences	Count	Percent
0	115	30.75
1	3	0.80
2	62	16.58
3	7	1.87
4	51	13.64
5	4	1.07
6	28	7.49
7	7	1.87
8	19	5.08
9	3	0.80
10	16	4.28
11	2	0.53
12	12	3.21
13	3	0.80
14	10	2.67
15	2	0.53
16	5	1.34
17	1	0.27
18	5	1.34
19	1	0.27
20	4	1.07
21	1	0.27
22	2	0.53
23	1	0.27
24	1	0.27
25	1	0.27
26	1	0.27
28	1	0.27
30	1	0.27
38	1	0.27
40	1	0.27
54	1	0.27
56	1	0.27
75	1	0.27
N=	374	
*=	21	

Absences:  $21/395 = 5.32\%$

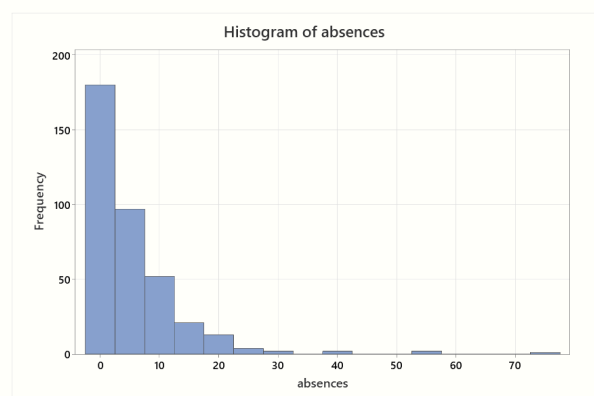
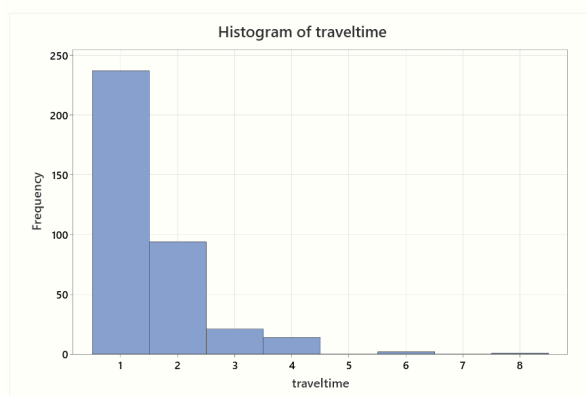
## 2. Identificar el mecanismo que ocasiona datos faltantes (MCAR, MAR, NMAR)}



Traveltime no tiene una correlación mayor a 0.12, mientras que absences llega hasta 0.17.

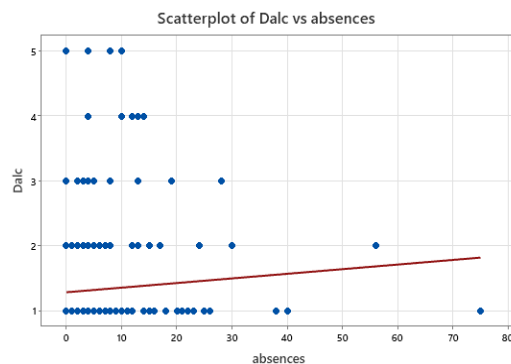
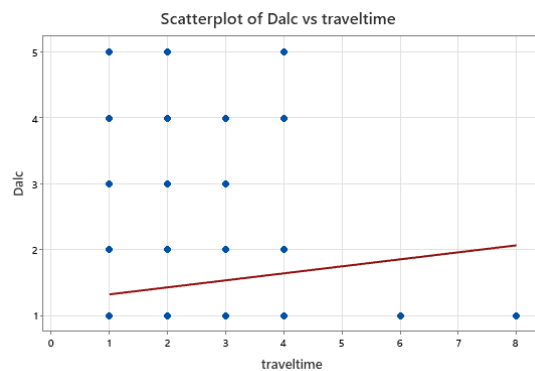
Por lo que parece ser *MCAR*

## 3. Obtener estadísticas descriptivas de los datos (histograma, media, desviación estándar, mediana, moda, etc).



## Statistics

Variable	N	N*	Mean	SE Mean	StDev	Minimum	Q1	Median	Q3	Maximum
traveltime	369	26	1.5285	0.0470	0.9028	1.0000	1.0000	1.0000	2.0000	8.0000
absences	374	21	5.543	0.418	8.089	0.000	0.000	3.500	8.000	75.000



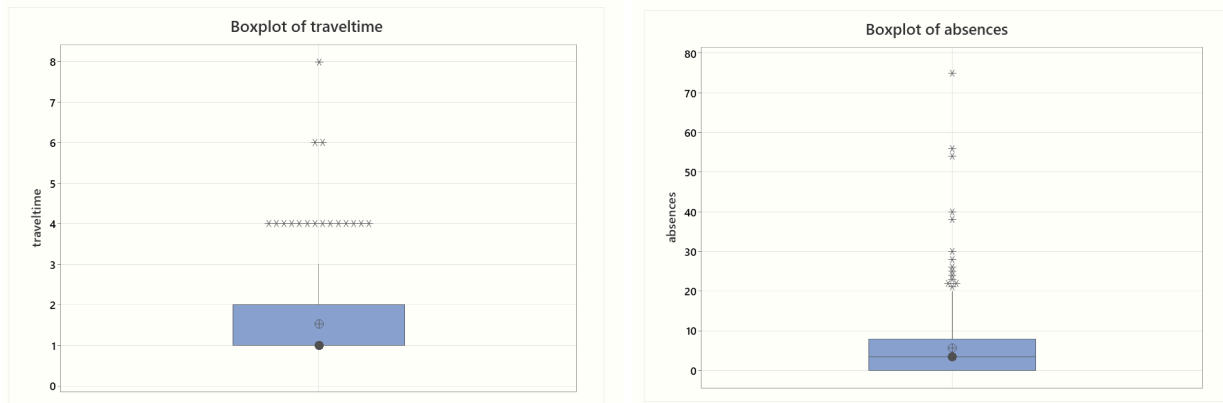
## 4.Utilizar el método de imputación adecuado para cada una de las variables con datos faltantes.

◦Imputación Simple: Media, Mediana, Moda

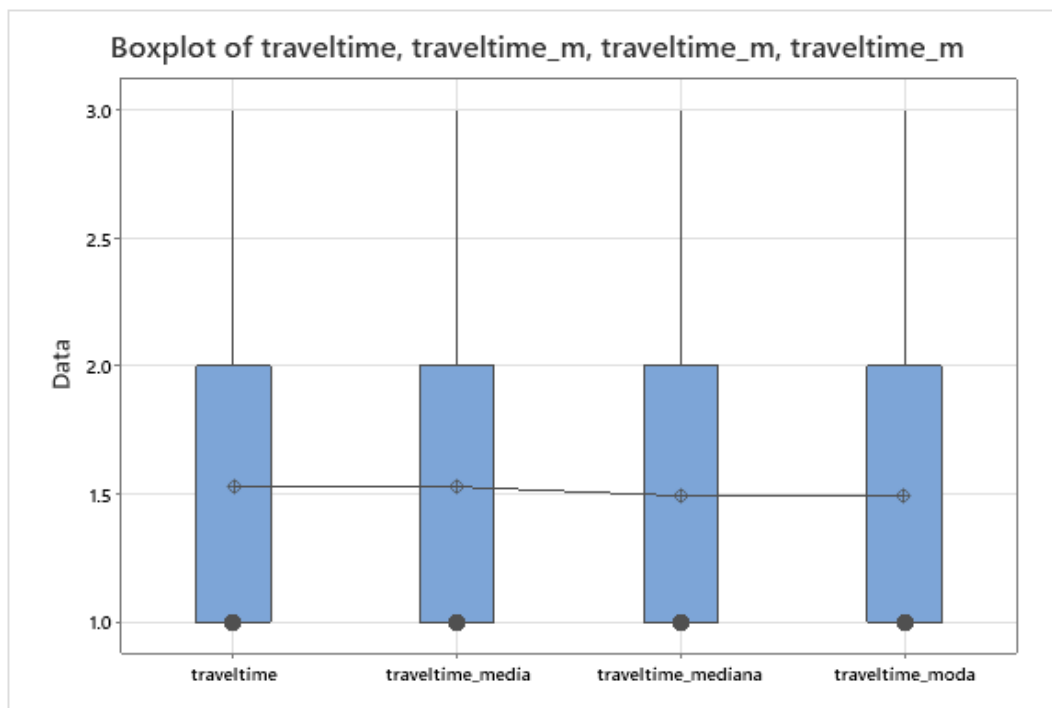
C13	C14	C15	C16	C34	C35	C36
traveltime	traveltime_media	traveltime_mediana	traveltime_moda	absences_mean	absences_mediana	absences_mode
1	1.00000	1	1	15.0000	15.0	15
1	1.00000	1	1	12.0000	12.0	12
2	2.00000	2	2	2.0000	2.0	2
1	1.00000	1	1	5.5428	3.5	0
1	1.00000	1	1	13.0000	13.0	13
2	2.00000	2	2	3.0000	3.0	3
1	1.00000	1	1	4.0000	4.0	4
2	2.00000	2	2	2.0000	2.0	2
1	1.00000	1	1	0.0000	0.0	0
1	1.00000	1	1	2.0000	2.0	2
2	2.00000	2	2	0.0000	0.0	0
1	1.00000	1	1	0.0000	0.0	0
1	1.00000	1	1	16.0000	16.0	16
1	1.00000	1	1	10.0000	10.0	10
1	1.00000	1	1	2.0000	2.0	2
4	4.00000	4	4	14.0000	14.0	14
..	1.52846	1	1	10.0000	10.0	10
1	1.00000	1	1	14.0000	14.0	14
2	2.00000	2	2	4.0000	4.0	4
1	1.00000	1	1	14.0000	14.0	14
1	1.00000	1	1	2.0000	2.0	2
..	1.52846	1	1	18.0000	18.0	18
2	2.00000	2	2	10.0000	10.0	10

## 5. Realizar un boxplot e interpretarlo.

*Box Plots individuales sin imputación con outliers:*



*Traveltime, con imputaciones, sin outliers:*

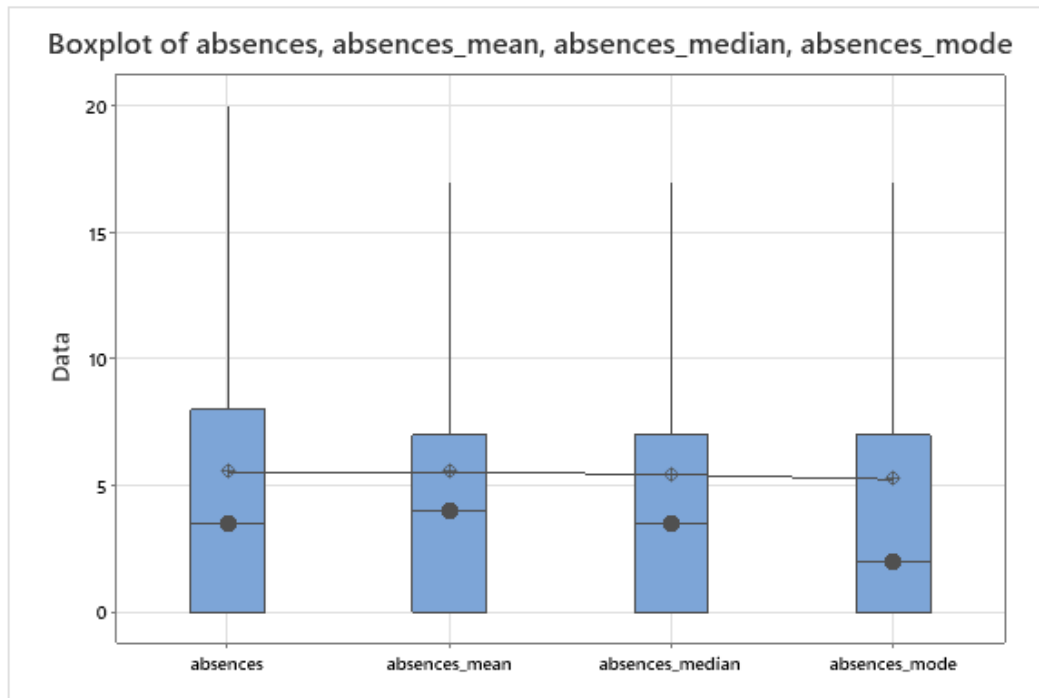


⊕ - Promedio

● - Mediana

Incluso con las imputaciones la mediana media y moda son bastantes similares, y esto era de esperarse dado que la mediana y moda son 1, y el promedio 1.5, por lo que al imputar no tuvieron un impacto tan grande en los box plots.

*Absences, con imputaciones, sin outliers:*

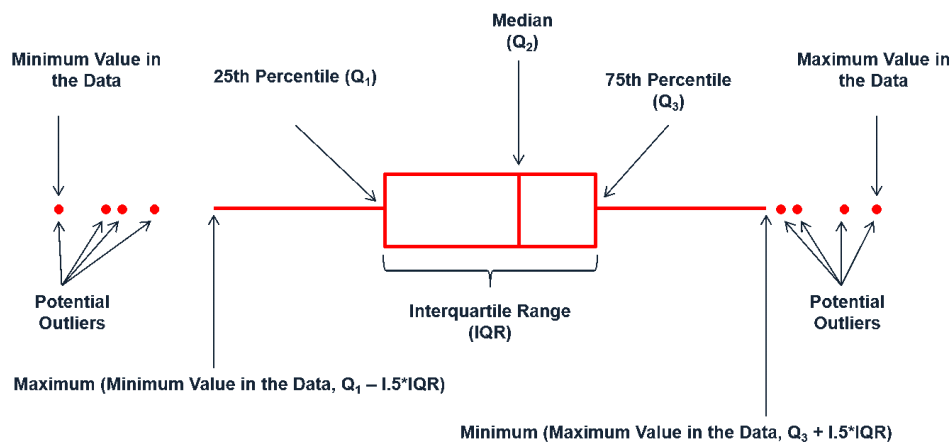


⊕ - Promedio

● - Mediana

Contamos con un rango intercuartil, máximos y mínimos similares entre el promedio, mediana y moda, sin embargo la mediana es la que más parece fluctuar. De la misma manera los promedios parecen ser similares entre todos pero disminuyendo ligeramente.

## ¿Cómo leer boxplots?



**Extra:**

Outliers! Porque decidí ignorarlos en los boxplots (para hacerlos más legibles)

