

Anàlisi descriptiva

Anna Salazar

2022-07-05



UNIVERSITAT POLITÈCNICA DE CATALUNYA
BARCELONATECH

Facultat de Matemàtiques i Estadística



UNIVERSITAT_{DE}
BARCELONA

Contents

Descripció de la base de dades	1
Anàlisi univariant	2
Variables categòriques	2
Tractament de les dades mancants	2
Variables numèriques	2
Variables qualitatives	2
Variables numèriques	10

Descripció de la base de dades

La base de dades està formada per les següents tres taules:

accident és un llistat d'accidents de tràfic ocorreguts al desembre de 2015 als Estats Units:

- ST_CASE: codi de l'accident (PK)
- DAY: dia de l'accident (de l'1 al 31)
- HOUR: hora de l'accident (99 = desconeguda)
- MINUTE: minut de l'accident (99 = desconegut)
- RUR_URB: informació sobre la localització (1 = Rural, 2 = Urbà, 6 = Via no classificada, 8 = No registrat, 9 = Desconegut)
- DAY_WEEK: dia de la setmana (1 = Diumenge, 2 = Dilluns, ..., 7 = Dissabte)
- FATALS: nombre de ferits a l'accident
- DRUNK_DR: nombre de conductors beguts involucrats a l'accident

person és un llistat de totes les persones (conductors, passatgers o vianants) involucrades als accidents

- ST_CASE: codi de l'accident al qual està involucrada la persona (PK)
- PER_NO: número de persona dins de cada accident (PK)
- AGE: edat de la persona (998 = No registrada, 999 = Desconeguda)
- SEX: sexe de la persona (1 = home, 2 = dona, 8 = No registrat, 9 = Desconegut)
- PER_TYP: tipus de persona (1 = conductor, 2 = ocupant, resta de codis = altres)
- DOA: tipus de víctima (0 = sobrevisant, 7 = mort a l'accident, 8 = mort al trasllat, 9 = Desconegut)

vehicle és un llistat de tots els vehicles involucrats als accidents

- ST_CASE: codi de l'accident al qual està involucrat el vehicle (PK)
- VEH_NO: número de vehicle dins de cada accident (PK)
- HIT_RUN: identificador de vehicle fugit (0 = No, 1 = Sí, 9 = Desconegut)
- TRAV_SP: velocitat estimada (mph)* del vehicle quan va tenir l'accident (997,998 i 999 = Desconegut)
- PREV_SP: indicador d'existència de límit de velocitat permesa just abans de l'accident (997,998 i 999 = Desconegut)

*mph vol dir milles per hora. 100 mph=161 Km/h

Anàlisi univariant

Variables categòriques

Tractament de les dades mancants

Variables numèriques

[1] 60.89408

[1] 3.01731

[1] 0.5081785

[1] 0.4367159

Variables qualitatives

Dia de l'accident

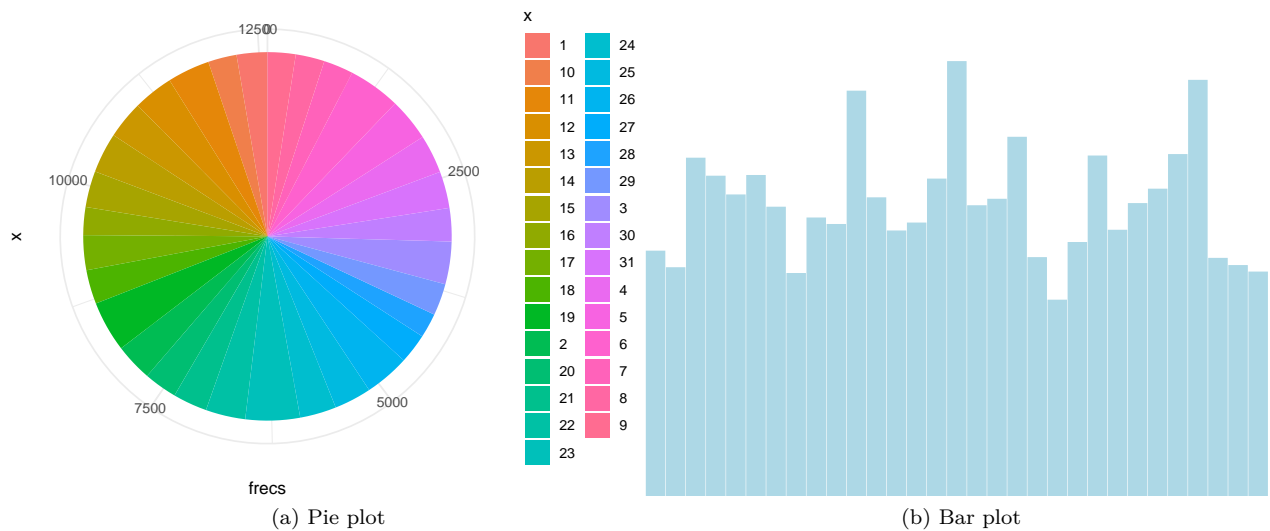


Figure 1: Anàlisi descriptiu de la variable Dia

Hora de l'accident

Table 1: Resum numèric de la variable Hora de l'accident

N.Valid	Min	Q1	Median	Mean	Std.Dev	Q3	Max	IQR
12539	0	9	15	13.51982	6.279563	18	23	9

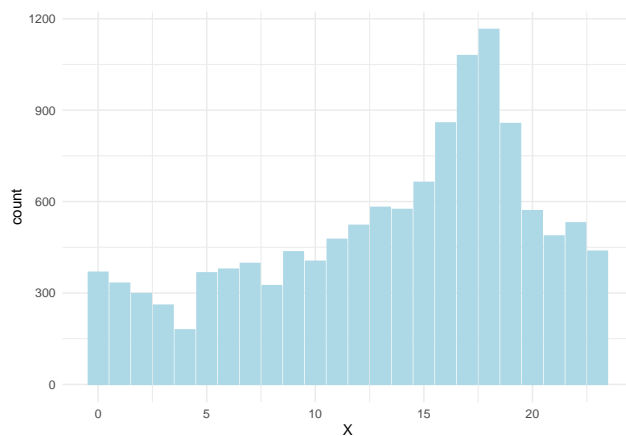


Figure 2: Histograma de la variable Hora de l'accident

Minut de l'accident

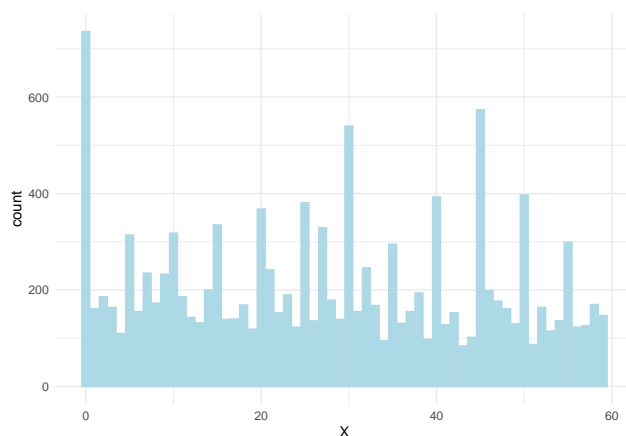


Figure 3: Histograma de la variable Minut de l'accident

Table 2: Resum numèric de la variable Minut de l'accident

N.Valid	Min	Q1	Median	Mean	Std.Dev	Q3	Max	IQR
12530	0	13	27	27.72674	17.36875	43	59	30

Localització

Variable	Stats / Values	Freqs (% of Valid)	Missing
Localització [factor]	1. Rural	5138 (40.8%)	0 (0.0%)
	2. Urbà	6193 (49.2%)	
	3. Via no classificada	37 (0.3%)	
	4. No registrat	1209 (9.6%)	
	5. Desconegut	17 (0.1%)	

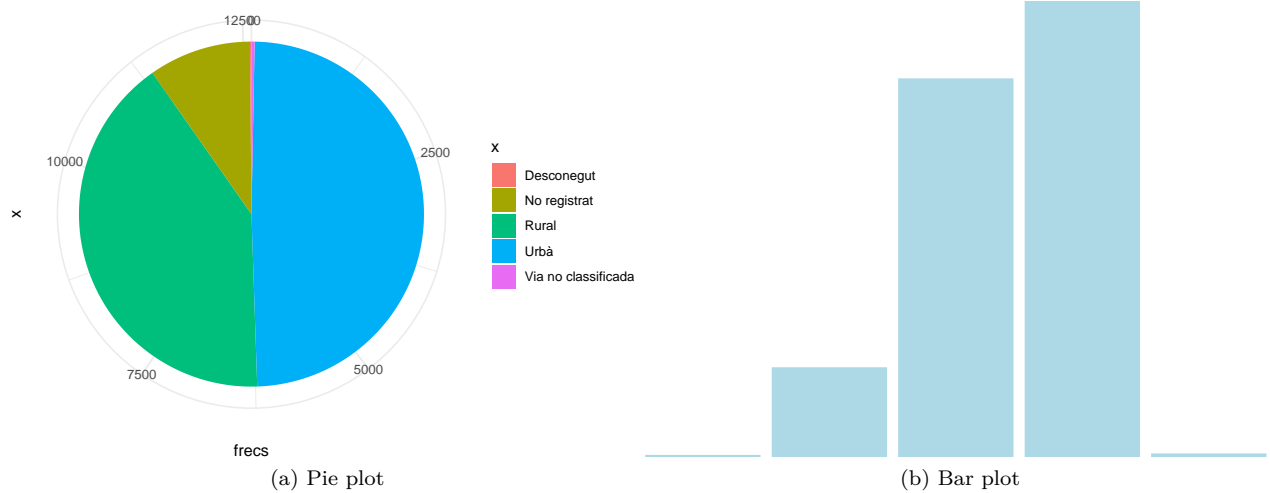


Figure 4: Anàlisi descriptiu de la variable Localització

Dia de la setmana

Variable	Stats / Values	Freqs (% of Valid)	Missing
Dia de la setmana [factor]	1. Diumenge	1690 (13.4%)	0 (0.0%)
	2. Dilluns	1422 (11.3%)	
	3. Dimarts	1848 (14.7%)	
	4. Dimecres	2001 (15.9%)	
	5. Dijous	1979 (15.7%)	
	6. Divendres	1680 (13.3%)	
	7. Dissabte	1974 (15.7%)	

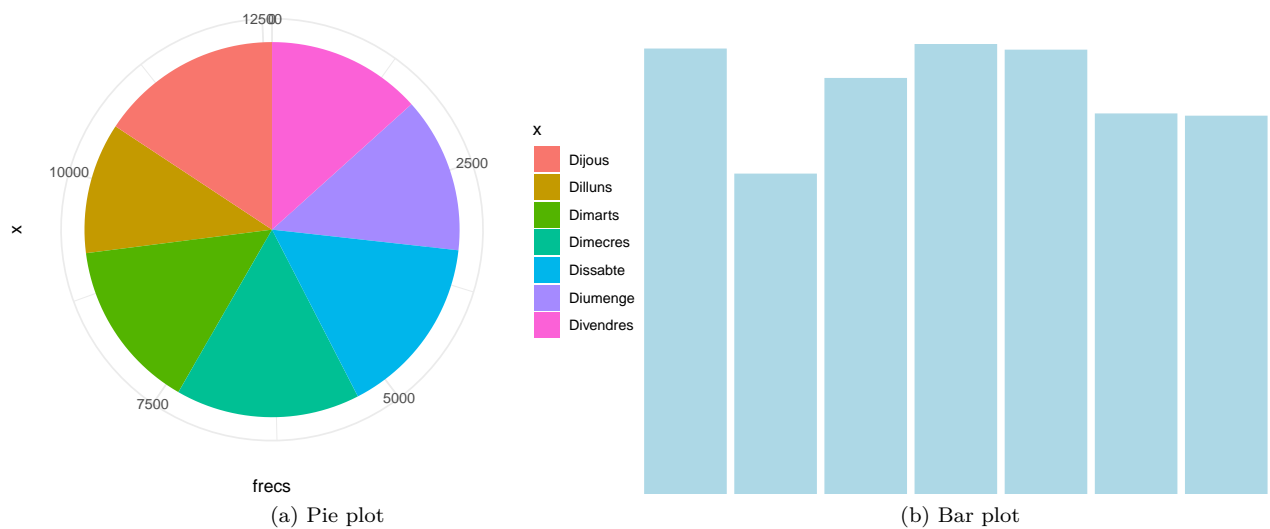


Figure 5: Anàlisi descriptiu de la variable Dia de la setmana

Nombre de ferits a l'accident

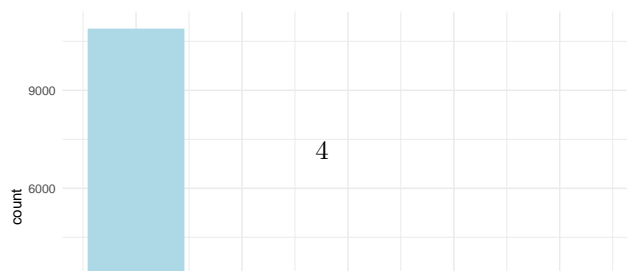


Table 5: Resum numèric de la variable Nombre de ferits a l'accident

N.Valid	Min	Q1	Median	Mean	Std.Dev	Q3	Max	IQR
12594	1	1	1	1.19287	0.5653175	1	5	0

Nombre de conductors beguts

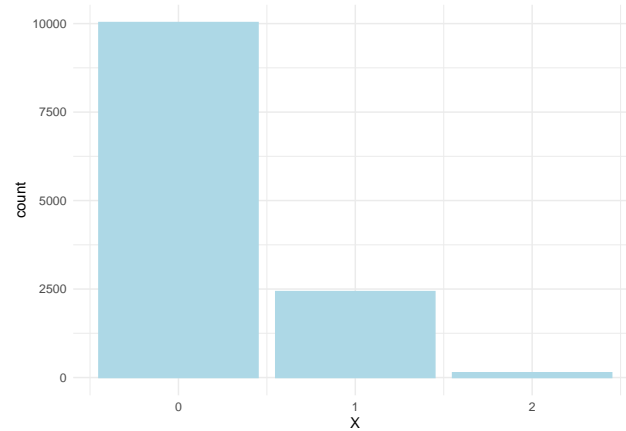


Figure 7: Histograma de la variable Nombre de conductors beguts

Table 6: Resum numèric de la variable Nombre de conductors beguts

N.Valid	Min	Q1	Median	Mean	Std.Dev	Q3	Max	IQR
12594	0	0	0	0.2143878	0.4359339	0	2	0

Nombre de persones implicades en l'accident

Table 7: Resum numèric de la variable Nombre de persones implicades en l'accident

N.Valid	Min	Q1	Median	Mean	Std.Dev	Q3	Max	IQR
12594	1	2	4	4.682071	5.237101	5	53	3

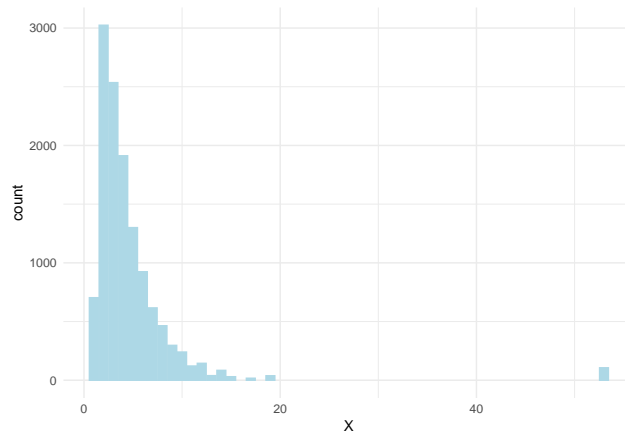


Figure 8: Anàlisi descriptiu de la variable Nombre de la persona

Edat

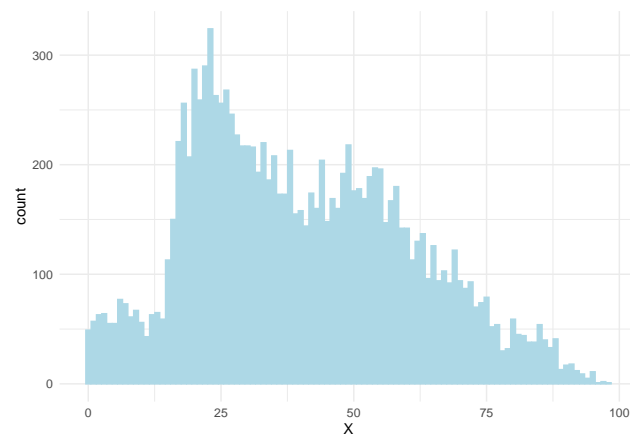


Figure 9: Histograma de la variable Edat

Table 8: Resum numèric de la variable Edat

N.Valid	Min	Q1	Median	Mean	Std.Dev	Q3	Max	IQR
12214	0	24	38	40.10553	20.44847	55	98	31

Sexe

Variable	Stats / Values	Freqs (% of Valid)	Missing
Sexe [factor]	1. Home	7960 (63.2%)	0 (0.0%)
	2. Dona	4327 (34.4%)	
	3. No registrat	149 (1.2%)	
	4. Desconegut	158 (1.3%)	

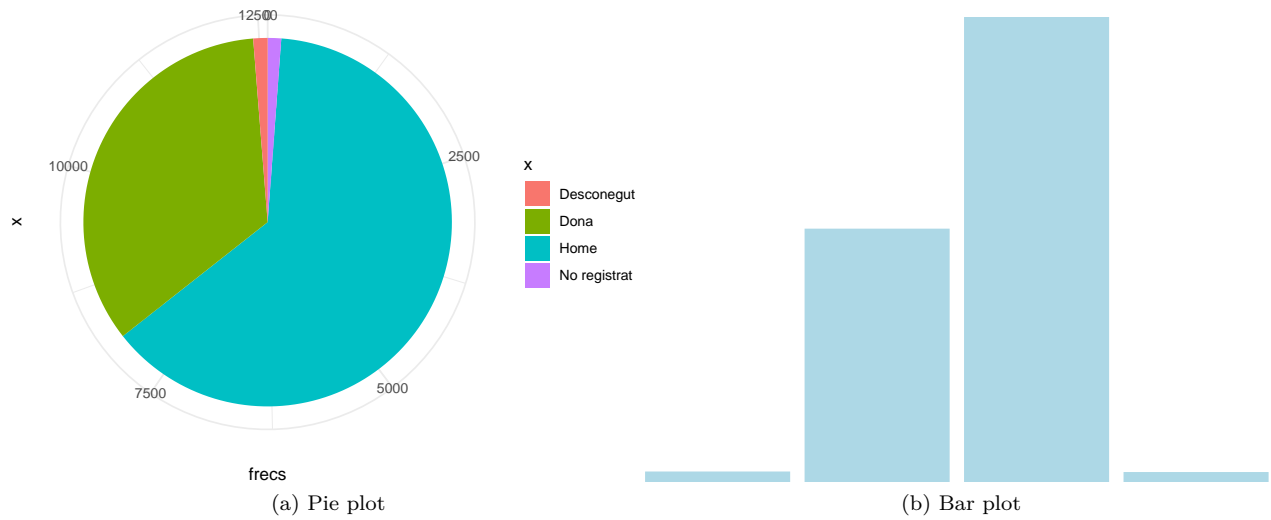


Figure 10: Anàlisi descriptiu de la variable Sexe

Tipus de persona

Variable	Stats / Values	Freqs (% of Valid)	Missing
Tipus de persona [factor]	1. Conductor	7620 (60.5%)	0
	2. Ocupant	4083 (32.4%)	(0.0%)
	3. Altres	891 (7.1%)	

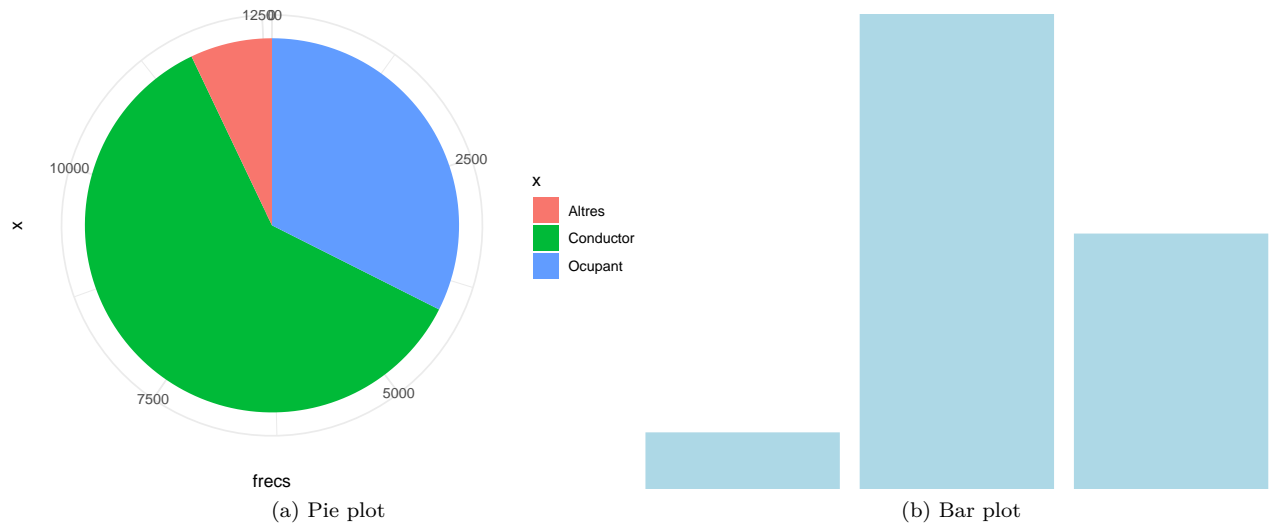


Figure 11: Anàlisi descriptiu de la variable Tipus de persona

Tipus de víctima

Variable	Stats / Values	Freqs (% of Valid)	Missing
Tipus de víctima [factor]	1. Sobrevisu	9798 (77.8%)	0
	2. Mort a l'accident	2752 (21.9%)	(0.0%)
	3. Mort al trasllat	41 (0.3%)	
	4. Desconegut	3 (0.0%)	

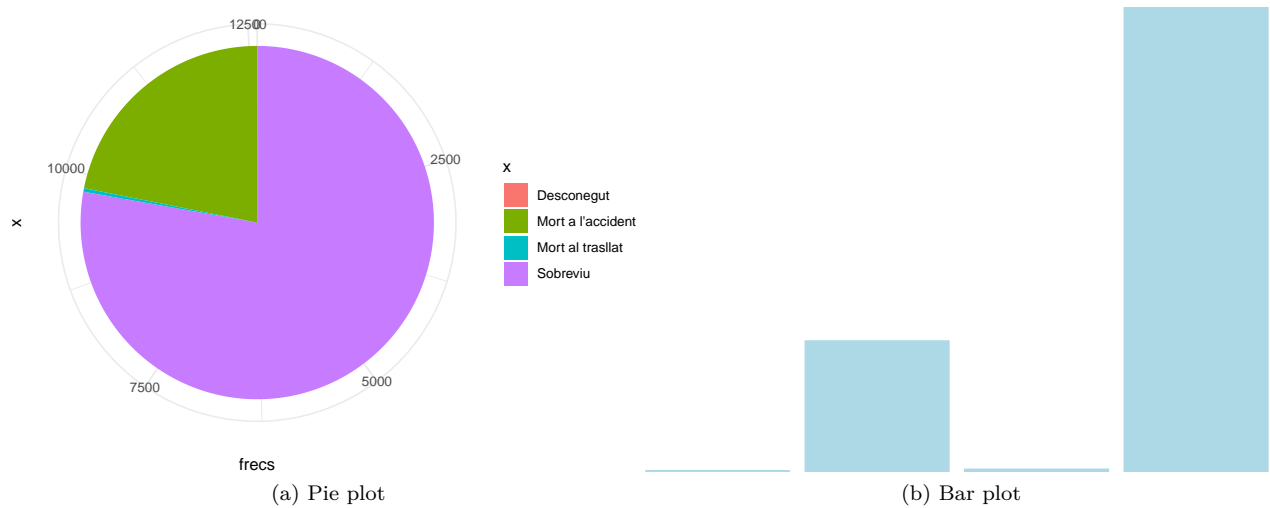


Figure 12: Anàlisi descriptiu de la variable Tipus de víctima

Nombre de vehicles involucrats en l'accident

Table 12: Resum numèric de la variable Nombre de vehicles involucrats en l'accident

N.Valid	Min	Q1	Median	Mean	Std.Dev	Q3	Max	IQR
12594	1	2	2	2.2074	1.081168	3	6	1

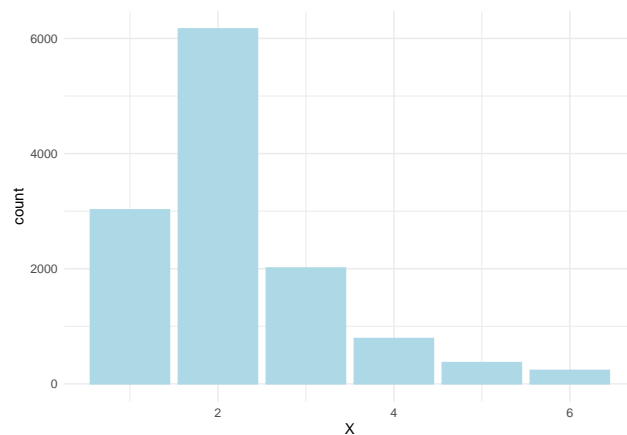


Figure 13: Anàlisi descriptiu de la variable Nombre de vehicles involucrats en l'accident

Identificació del vehicle

Variable	Stats / Values	Freqs (% of Valid)	Missing
Identificació del vehicle [factor]	1. No	12103 (96.1%)	0
	2. Sí	485 (3.9%)	(0.0%)
	3. Desconegut	6 (0.0%)	

Velocitat estimada

Table 14: Resum numèric de la variable Velocitat estimada

N.Valid	Min	Q1	Median	Mean	Std.Dev	Q3	Max	IQR
4925	0	25	46	42.4869	25.40908	60	134	35

Indicador d'existència de límit de velocitat permesa just abans de l'accident

Variable	Stats / Values	Freqs (% of Valid)	Missing
----------	----------------	--------------------	---------

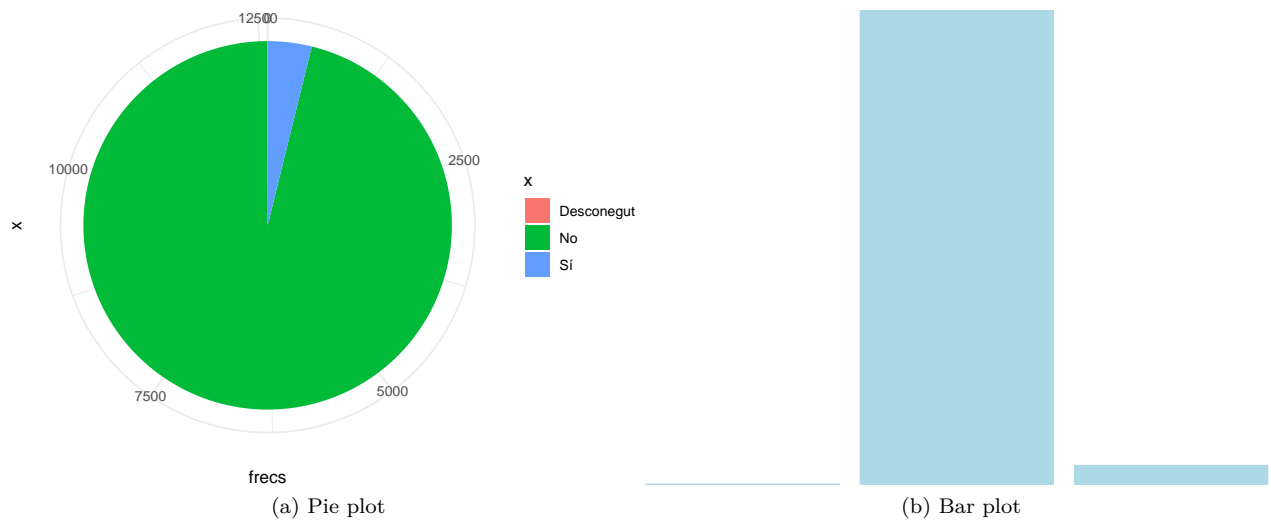


Figure 14: Anàlisi descriptiu de la variable Identificació del vehicle

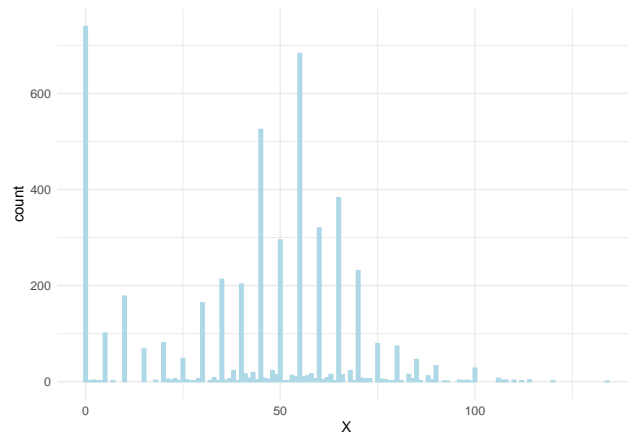


Figure 15: Histograma de la variable Velocitat estimada

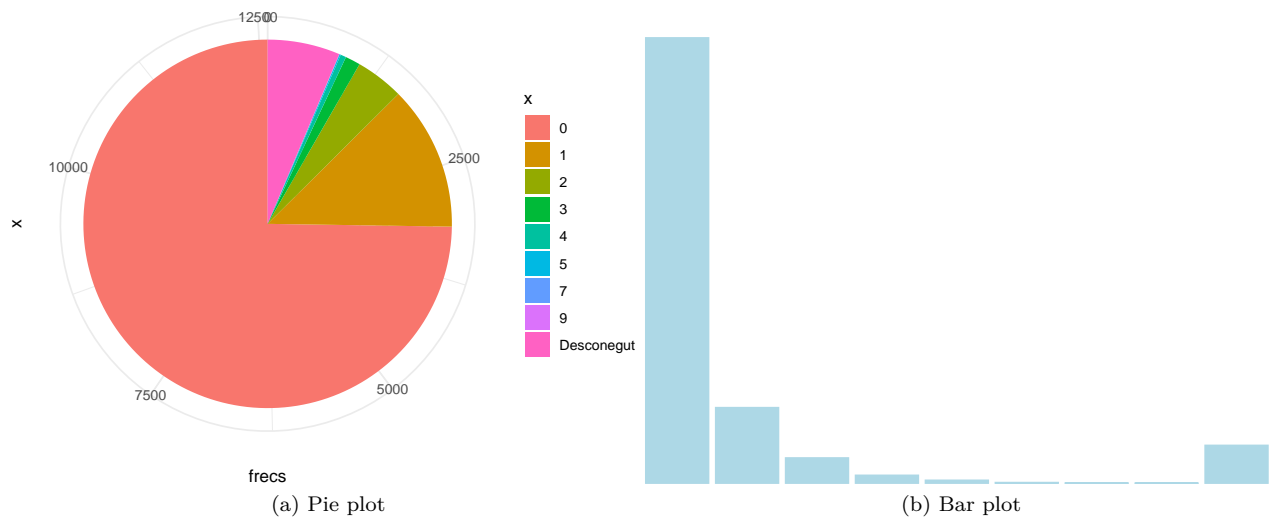


Figure 16: Anàlisi descriptiu de la variable Indicador d'existència de límit de velocitat