mpg

Autogenerated data summary from dataMaid

2020-06-07 19:08:42

Part 1

Data report overview

The dataset examined has the following dimensions:

Feature	Result
Number of observations	234
Number of variables	11

Checks performed

The following variable checks were performed, depending on the data type of each variable:

	character	factor	labelled	haven labelled	numeric	integer	logical	Date
Identify miscoded missing values	×	×	×	×	×	×		×
Identify prefixed and suffixed whitespace	×	×	×	×				
Identify levels with < 6 obs.	×	×	×	×				
Identify case issues	×	×	×	×				
Identify misclassified numeric or integer variables	×	×	×	×				
Identify outliers					×	×		×

Please note that all numerical values in the following have been rounded to 2 decimals.

Part 2

Summary table

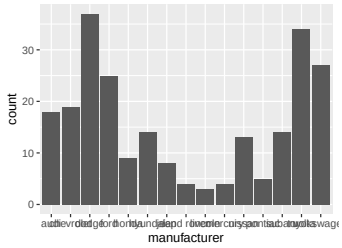
	Variable class	# unique values	Missing observations	Any problems?
manufacturer	character	15	0.00 %	×
model	character	38	0.00 %	×
displ	numeric	35	0.00 %	
year	integer	2	0.00 %	
cyl	integer	4	0.00 %	×
trans	character	10	0.00 %	×
drv	character	3	0.00 %	
cty	integer	21	0.00 %	×
hwy	integer	27	0.00 %	×
fl	character	5	0.00 %	×
class	character	7	0.00 %	×

Part 3

Variable list

manufacturer

Feature	Result
Variable type	character
Number of missing obs.	0 (0 %)
Number of unique values	15
Mode	“dodge”

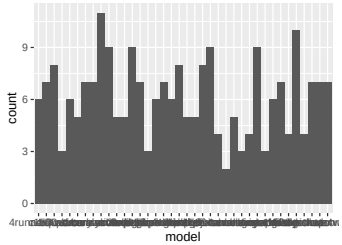


- Note that the following levels have at most five observations: "land rover", "lincoln", "mercury", "pontiac".

---

model

Feature	Result
Variable type	character
Number of missing obs.	0 (0 %)
Number of unique values	38
Mode	“caravan 2wd”

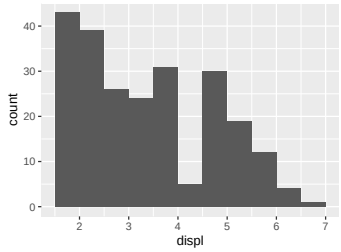


- Note that the following levels have at most five observations: "a6 quattro", "c1500 suburban 2wd", "corolla", "corvette", "expedition 2wd", "grand prix", "gti", "k1500 tahoe 4wd", "land cruiser wagon 4wd", "malibu" (5 additional values omitted).

---

displ

Feature	Result
Variable type	numeric
Number of missing obs.	0 (0 %)
Number of unique values	35
Median	3.3
1st and 3rd quartiles	2.4; 4.6
Min. and max.	1.6; 7

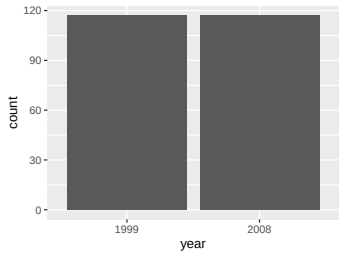


---

year

- Note that this variable is treated as a factor variable below, as it only takes a few unique values.

Feature	Result
Variable type	integer
Number of missing obs.	0 (0 %)
Number of unique values	2
Mode	“1999”
Reference category	1999

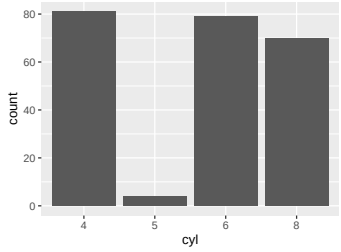


---

cyl

- Note that this variable is treated as a factor variable below, as it only takes a few unique values.

Feature	Result
Variable type	integer
Number of missing obs.	0 (0 %)
Number of unique values	4
Mode	“4”
Reference category	4

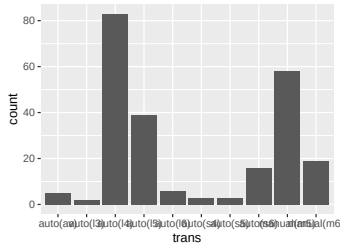


- The following suspected missing value codes enter as regular values: "8".
- Note that the following levels have at most five observations: "5".

---

trans

Feature	Result
Variable type	character
Number of missing obs.	0 (0 %)
Number of unique values	10
Mode	“auto(l4)”

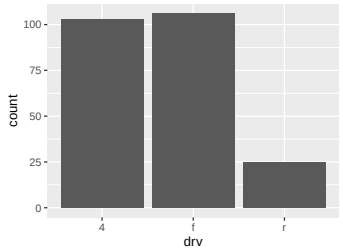


- Note that the following levels have at most five observations: "auto(av)", "auto(l3)", "auto(s4)", "auto(s5)".

---

drv

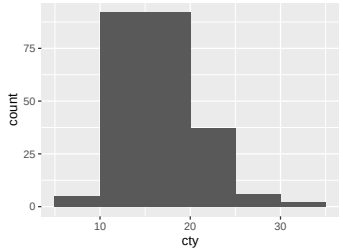
Feature	Result
Variable type	character
Number of missing obs.	0 (0 %)
Number of unique values	3
Mode	“f”



---

cty

Feature	Result
Variable type	integer
Number of missing obs.	0 (0 %)
Number of unique values	21
Median	17
1st and 3rd quartiles	14; 19
Min. and max.	9; 35

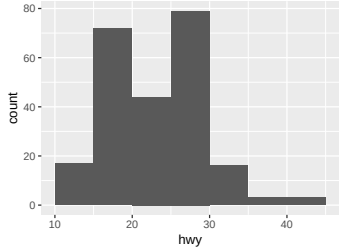


- Note that the following possible outlier values were detected: "28", "29", "33", "35".

---

hwy

Feature	Result
Variable type	integer
Number of missing obs.	0 (0 %)
Number of unique values	27
Median	24
1st and 3rd quartiles	18; 27
Min. and max.	12; 44

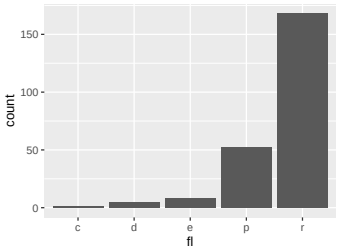


- Note that the following possible outlier values were detected: "32", "33", "34", "35", "36", "37", "41", "44".

---

fl

Feature	Result
Variable type	character
Number of missing obs.	0 (0 %)
Number of unique values	5
Mode	“r”

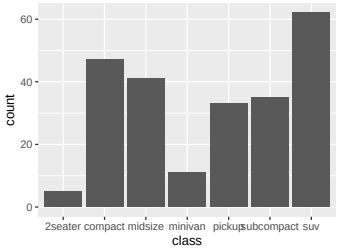


- Note that the following levels have at most five observations: "c", "d".

---

class

Feature	Result
Variable type	character
Number of missing obs.	0 (0 %)
Number of unique values	7
Mode	“suv”



- Note that the following levels have at most five observations: "2seater".

Report generation information:

- Created by: Could not determine from system (username: `esuess`).
- Report creation time: Sun Jun 07 2020 19:08:42
- Report was run from directory: `/home/esuess/Documents/Stat650/nycflights13`
- dataMaid v1.4.0 [Pkg: 2019-12-10 from CRAN (R 3.6.2)]
- R version 3.6.3 (2020-02-29).
- Platform: x86_64-pc-linux-gnu (64-bit)(Ubuntu 18.04.4 LTS).
- Function call: `makeDataReport(data = mpg)`