

Data Exploration: SELECTING CASES

SELECTING CASES

- datasets: built in - **datasets::mtcars**

Statistics summary

- Selecting particular cases from data set

Practical:

- Use selected cases in analysis
 - on the spot (see in console)
 - or, save into new data frame - for further analysis
 - format: `[dataset$var boolean rows, columns]`.
- Great way to focus analysis - by drilling down to subsets
 - → see if patterns emerge at different levels of abstraction

Pseudocode:

- Selecting dataset subgroup
 - Select a variable (`dataset-name$variable-name`)
 - Select specific values in variable (logical equivalence `==` or `>=` or `<=`)
- Doing something with this new subgroup
 - Creating a new data frame from this subgroup; or
 - Performing a function on the subgroup

SCRIPTS SUMMARY

LOAD DATA

```
?mtcars
data(mtcars)
mtcars
```

SELECTING CASES IN DATAFRAME

Perform a function on subset

```
mean(mtcars$qsec)
mean(mtcars$qsec[mtcars$cyl == 8])
median(mtcars$hp)
mean(mtcars$mpg[mtcars$hp > median(mtcars$hp)])
```

Create new dataframe from subset

```
cyl.8 <- mtcars[mtcars$cyl == 8, ]
mtcars[mtcars$cyl == 8 & mtcars$carb >= 4, ]
```

CLEAN UP

```
rm(list = ls())
```

Data Analysis Findings**1974 Motor Trends Magazine car tests
datasets::mtcars**

1974 Motor Trends Magazine car tests is a collection of 32 different cars, with information on each model's mpg, cylinders, displacement, hp, drive train ratio, weight, stop times, carburetor barrels, gears.

We first explore the amount of time it takes for the car to stop. For all cars in this dataset, the **mean** quarter-mile time for all cars is very slow at **17.85** seconds. You probably can walk a quarter mile that fast. The **mean** quarter-mile time for 8-cylinder cars is one second faster at **16.77** seconds.

Now we look at the horsepower of the cars. The **median** horsepower for all cars is **123** which is respectable. The **mean** MPG for cars above **median** horsepower is **15.41** miles per gallon.

Now we look at 8-cylinder cars. There are 14 models, of which 7 models have 4 or more barrel carburetors. Only one model, the Maserato Bora, has more than 4 barrels, with 8 barrels.

SCRIPTS & NOTES

- When prepare data for analysis → usually don't want every single case in dataset
 - → select **cases** - focus on **subset**

LOAD DATA

?mtcars #1974 Motor Trends Magazine car tests

data(mtcars)

- → workspace: 32 obs. in 11 variables

mtcars

```
> mtcars
```

	mpg	cyl	disp	hp	drat	wt	qsec	vs	am	gear	carb
Mazda RX4	21.0	6	160.0	110	3.90	2.620	16.46	0	1	4	4
Mazda RX4 Wag	21.0	6	160.0	110	3.90	2.875	17.02	0	1	4	4
Datsun 710	22.8	4	108.0	93	3.85	2.320	18.61	1	1	4	1
Hornet 4 Drive	21.4	6	258.0	110	3.08	3.215	19.44	1	0	3	1
Hornet Sportabout	18.7	8	360.0	175	3.15	3.440	17.02	0	0	3	2
Valiant	18.1	6	225.0	105	2.76	3.460	20.22	1	0	3	1
Duster 360	14.3	8	360.0	245	3.21	3.570	15.84	0	0	3	4
Merc 240D	24.4	4	146.7	62	3.69	3.190	20.00	1	0	4	2
Merc 230	22.8	4	140.8	95	3.92	3.150	22.90	1	0	4	2
Merc 280	19.2	6	167.6	123	3.92	3.440	18.30	1	0	4	4
Merc 280C	17.8	6	167.6	123	3.92	3.440	18.90	1	0	4	4
Merc 450SE	16.4	8	275.8	180	3.07	4.070	17.40	0	0	3	3
Merc 450SL	17.3	8	275.8	180	3.07	3.730	17.60	0	0	3	3
Merc 450SLC	15.2	8	275.8	180	3.07	3.780	18.00	0	0	3	3
Cadillac Fleetwood	10.4	8	472.0	205	2.93	5.250	17.98	0	0	3	4
Lincoln Continental	10.4	8	460.0	215	3.00	5.424	17.82	0	0	3	4
Chrysler Imperial	14.7	8	440.0	230	3.23	5.345	17.42	0	0	3	4
Fiat 128	32.4	4	78.7	66	4.08	2.200	19.47	1	1	4	1
Honda Civic	30.4	4	75.7	52	4.93	1.615	18.52	1	1	4	2
Toyota Corolla	33.9	4	71.1	65	4.22	1.835	19.90	1	1	4	1
Toyota Corona	21.5	4	120.1	97	3.70	2.465	20.01	1	0	3	1
Dodge Challenger	15.5	8	318.0	150	2.76	3.520	16.87	0	0	3	2
AMC Javelin	15.2	8	304.0	150	3.15	3.435	17.30	0	0	3	2
Camaro Z28	13.3	8	350.0	245	3.73	3.840	15.41	0	0	3	4
Pontiac Firebird	19.2	8	400.0	175	3.08	3.845	17.05	0	0	3	2
Fiat X1-9	27.3	4	79.0	66	4.08	1.935	18.90	1	1	4	1
Porsche 914-2	26.0	4	120.3	91	4.43	2.140	16.70	0	1	5	2
Lotus Europa	30.4	4	95.1	113	3.77	1.513	16.90	1	1	5	2
Ford Pantera L	15.8	8	351.0	264	4.22	3.170	14.50	0	1	5	4
Ferrari Dino	19.7	6	145.0	175	3.62	2.770	15.50	0	1	5	6
Maserati Bora	15.0	8	301.0	335	3.54	3.570	14.60	0	1	5	8
Volvo 142E	21.4	4	121.0	109	4.11	2.780	18.60	1	1	4	2

- Have 32 different cars
- Info on mpg | cylinders | displacement | hp | drive train ratio | weight ... carburetor barrels | gears

SELECTING CASES IN DATAFRAME**mean quarter-mile time** (for all cars)`mean(mtcars$qsec)``#[1] 17.84875` # very slow - probably can walk a quarter mile that fast**mean quarter-mile time** (for 8-cylinder cars)

- use square brackets to indicate what to select - in this format: [rows]
- == equivalence sign - boolean (= sign is an assignment operator like <-)

`mean(mtcars$qsec[mtcars$cyl == 8])`

- want mean
- data set is mtcars
- look at just \$qsec variable (same as in mean above)
- for subset → square brackets - row # want to select - in variable \$cyl

`#[1] 16.77214` # 1 second faster**median horsepower** (for all cars)`median(mtcars$hp)``#[1] 123` # respectable**mean MPG for cars above median horsepower**`mean(mtcars$mpg[mtcars$hp > median(mtcars$hp)])` # long bc specify dataset w/every variable

- want mean
- of mpg variable - in dataset mtcars
- square brackets to indicate rows selecting
 - selecting any car whose hp - is above median value of hp

`#[1] 15.40667`**Create new data frame for 8-cylinder cars**

- To create a new data frame, tell R which rows and columns to copy
 - format: `[dataset$var boolean rows, columns]`.

`cyl.8 <- mtcars[mtcars$cyl == 8, ]`

- come from mtcars
- only cases - where variable cyl (cylinders) within that dataset - equal to 8
- → workspace: Data cyl.8 14 obs. of 11 variables

	mpg	cyl	disp	hp	drat	wt	qsec	vs	am	gear	carb
Hornet Sportabout	18.7	8	360.0	175	3.15	3.440	17.02	0	0	3	2
Duster 360	14.3	8	360.0	245	3.21	3.570	15.84	0	0	3	4
Merc 450SE	16.4	8	275.8	180	3.07	4.070	17.40	0	0	3	3
Merc 450SL	17.3	8	275.8	180	3.07	3.730	17.60	0	0	3	3
Merc 450SLC	15.2	8	275.8	180	3.07	3.780	18.00	0	0	3	3
Cadillac Fleetwood	10.4	8	472.0	205	2.93	5.250	17.98	0	0	3	4
Lincoln Continental	10.4	8	460.0	215	3.00	5.424	17.82	0	0	3	4
Chrysler Imperial	14.7	8	440.0	230	3.23	5.345	17.42	0	0	3	4
Dodge Challenger	15.5	8	318.0	150	2.76	3.520	16.87	0	0	3	2
AMC Javelin	15.2	8	304.0	150	3.15	3.435	17.30	0	0	3	2
Camaro Z28	13.3	8	350.0	245	3.73	3.840	15.41	0	0	3	4
Pontiac Firebird	19.2	8	400.0	175	3.08	3.845	17.05	0	0	3	2
Ford Pantera L	15.8	8	351.0	264	4.22	3.170	14.50	0	1	5	4
Maserati Bora	15.0	8	301.0	335	3.54	3.570	14.60	0	1	5	8

- spreadsheet format results:
 - see Hornet, Duster → tells you are early 70s
 - see Javelin, Firebird, Pantera, Maseroti Bora

Select 8-cylinder cars with 4+ barrel carburetors

```
mtcars[mtcars$cyl == 8 & mtcars$carb >= 4, ]
```

- combine conditions with &
- → this time print out in console (not into new object) - to see which cars meet the criteria

```
> mtcars[mtcars$cyl == 8 & mtcars$carb >= 4, ]
```

	mpg	cyl	disp	hp	drat	wt	qsec	vs	am	gear	carb
Duster 360	14.3	8	360	245	3.21	3.570	15.84	0	0	3	4
Cadillac Fleetwood	10.4	8	472	205	2.93	5.250	17.98	0	0	3	4
Lincoln Continental	10.4	8	460	215	3.00	5.424	17.82	0	0	3	4
Chrysler Imperial	14.7	8	440	230	3.23	5.345	17.42	0	0	3	4
Camaro Z28	13.3	8	350	245	3.73	3.840	15.41	0	0	3	4
Ford Pantera L	15.8	8	351	264	4.22	3.170	14.50	0	1	5	4
Maserati Bora	15.0	8	301	335	3.54	3.570	14.60	0	1	5	8

- results: list of 8-cyl cars that have 4 or more barrels
 - only 1 has > 4 , Maserati Bora has 8 barrels

CLEAN UP

```
rm(list = ls())
```

Analysis Results

1974 Motor Trends Magazine car tests is a collection of 32 different cars, with information on each model's mpg, cylinders, displacement, hp, drive train ratio, weight, stop times, carburetor barrels, gears.

We first explore the amount of time it takes for the car to stop. For all cars in this dataset, the **mean** quarter-mile time for all cars is very slow at **17.85** seconds. You probably can walk a quarter mile that fast. The **mean** quarter-mile time for 8-cylinder cars is one second faster at **16.77** seconds. Now we look at the horsepower of the cars. The **median** horsepower for all cars is **123** which is respectable. The **mean** MPG for cars above **median** horsepower is **15.41** miles per gallon.

Now we look at 8-cylinder cars. There are 14 models, of which 7 models have 4 or more barrel carburetors. Only 1 model, the Maserati Bora has more than 4 barrels, with 8 barrels.