

# PROJECT #1

Find some interesting data on the web

- something that challenges and interests you
- there are many data sources on the web
- use google and some imagination

Criteria for selection

- more than 500 data points (observations)
- more than 10 attributes
- the more the better (you can always reduce it)

Deliverables

- 2-page report that describes the data and justifies your choice
- a URL to the data source

Due date

- Tuesday, September 20, 11:59pm

# PROJECT #1: DATASET EXAMPLE

## Multivariate - Quantitative data and Categorical data

### Data Items

	A	B	C	D	E	F	G	H	I
1	Name	Country	Miles Per Gallon	Accceleration	Horsepower	weight	cylinders	year	price
2	Volkswagen Rabbit DI	Germany	43,1	21,5	48	1985	4	78	2400
3	Ford Fiesta	Germany	36,1	14,4	66	1800	4	78	1900
4	Mazda GLC Deluxe	Japan	32,8	19,4	52	1985	4	78	2200
5	Datsun B210 GX	Japan	39,4	18,6	70	2070	4	78	2725
6	Honda Civic CVCC	Japan	36,1	16,4	60	1800	4	78	2250
7	Oldsmobile Cutlass	USA	19,9	15,5	110	3365	8	78	3300
8	Dodge Diplomat	USA	19,4	13,2	140	3735	8	78	3125
9	Mercury Monarch	USA	20,2	12,8	139	3570	8	78	2850
10	Pontiac Phoenix	USA	19,2	19,2	105	3535	6	78	2800
11	Chevrolet Malibu	USA	20,5	18,2	95	3155	6	78	3275
12	Ford Fairmont A	USA	20,2	15,8	85	2965	6	78	2375
13	Ford Fairmont M	USA	25,1	15,4	88	2720	4	78	2275
14	Plymouth Volare	USA	20,5	17,2	100	3430	6	78	2700
15	AMC Concord	USA	19,4	17,2	90	3210	6	78	2300
16	Buick Century	USA	20,6	15,8	105	3380	6	78	3300
17	Mercury Zephyr	USA	20,8	16,7	85	3070	6	78	2425
18	Dodge Aspen	USA	18,6	18,7	110	3620	6	78	2700
19	AMC Concord D1	USA	18,1	15,1	120	3410	6	78	2425
20	Chevrolet MonteCarlo	USA	19,2	13,2	145	3425	8	78	3900
21	Buick RegalTurbo	USA	17,7	13,4	165	3445	6	78	4400
22	Ford Futura	Germany	18,1	11,2	139	3205	8	78	2525
23	Dodge Magnum XE	USA	17,5	13,7	140	4080	8	78	3000
24	Chevrolet Chevette	USA	30	16,5	68	2155	4	78	2100
25	Toyota Corona	Japan	27,5	14,2	95	2560	4	78	2975

### Data types

Quantitative (Numerical)

Categorical (Ordinal)

↑  
**Categorical**

↑  
**Quantitative**

↑  
Categorical (Ordinal)  
Quantitative

# PROJECT #1: NOTES ON DATASET

## Other data types are OK

- text, images, video, logs, etc.
- just convert them to numbers via appropriate mechanism as discussed in class
- must produce a spreadsheet of rows (data items) and attributes (columns)

## Categorical data

- color, brand, country, etc.
- convert into numbers by assigning a numerical ID

# QUESTIONS?

The course has been set up with Piazza

- <http://piazza.com/stonybrook/fall2016/cse332/home>
- please let me know if you cannot access it

Make use of this handy discussion forum

- ask questions of general interest
- give advice to peers (those who ask questions)
- give general feedback (observe etiquette)
- but obviously, don't provide actual solutions and aid in cheating