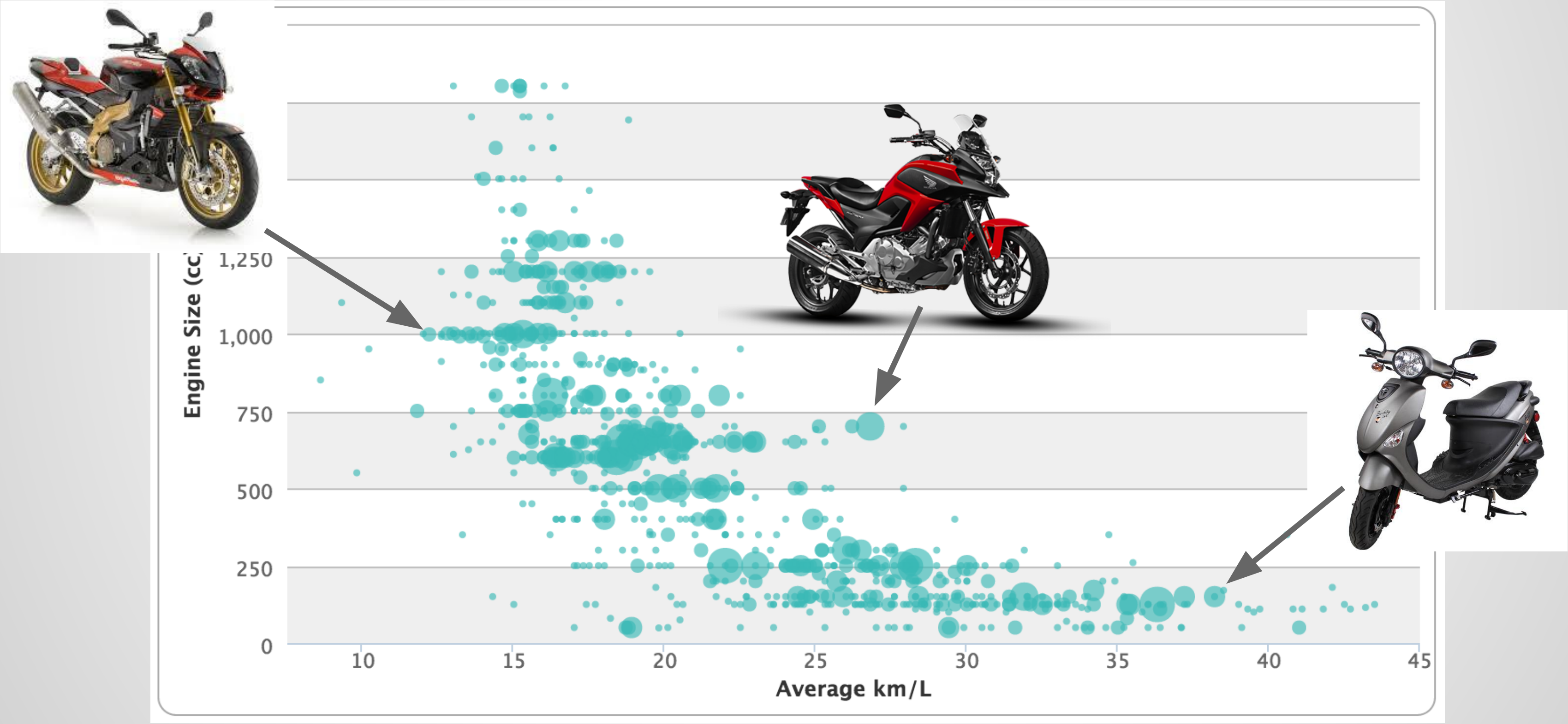
CSCI 4146 Data Science

Project Progress Presentation
By Derek Neil

Vehicle Fuel Economy



Data Source(s)

Real World Data

http://www.fuelly.com/motorcycle

http://www.fuelly.com/car

Institutional Data

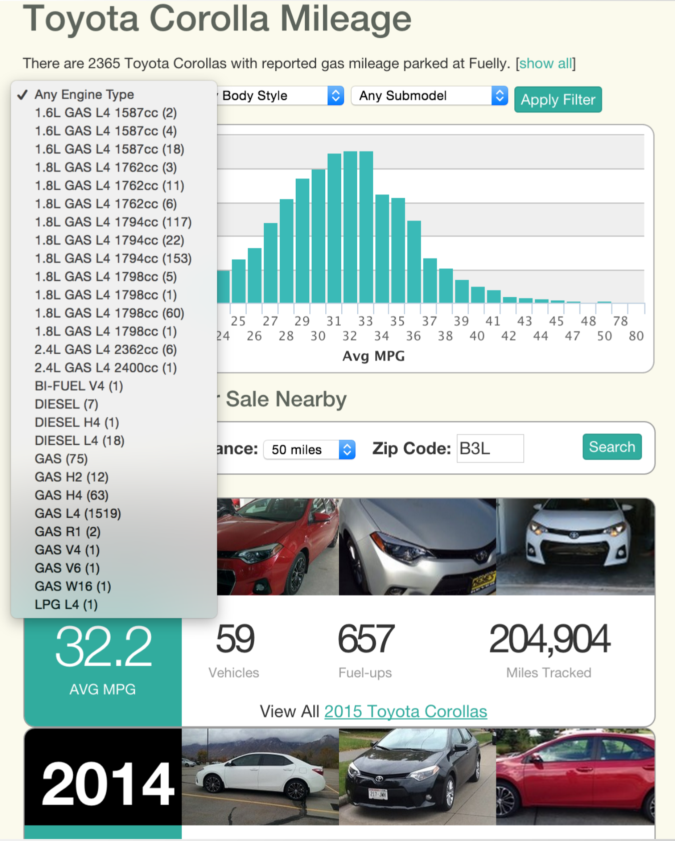
https://github.com/caesar0301/awesome-public-datasets#Transportation

https://www.fueleconomy.gov/feg/download.shtml

http://open.canada.ca/data/en/dataset/98f1a129-f628-4ce4-b24d-6f16bf24dd64

Real World Data (fuelly.com)

- Real World Conditions
- Large Sample Size*
- ~Normally Distributed
- Attributes
  - 175k cars over 20 years
  - 2431 make-models
  - Location Data*
  - Engine Size*
  - Assuming “Combined” City/Highway metric

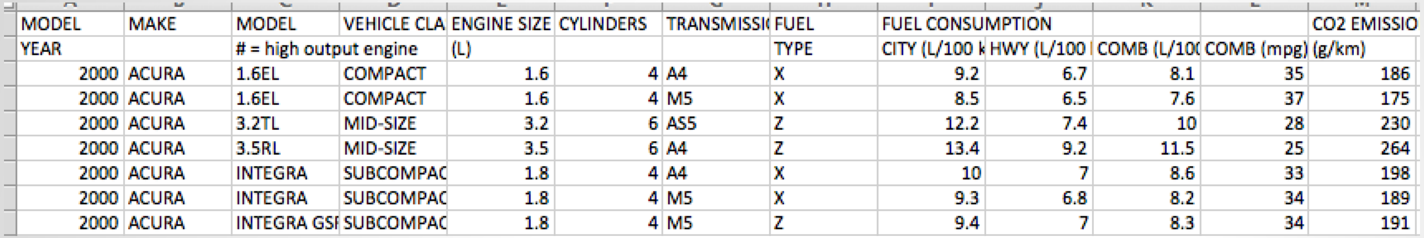


Toyota Corolla Mileage

There are 2365 Toyota Corollas with reported gas mileage parked at Fuelly. [show all]

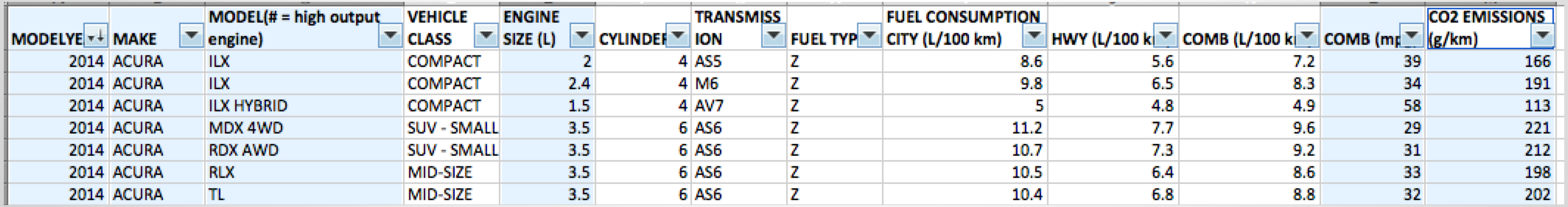
View All 2015 Toyota Corollas

Institutional Data (open.canada.ca)



MODEL	MAKE	MODEL	VEHICLE CLA	ENGINE SIZE	CYLINDERS	TRANSMISSI	FUEL	FUEL CONSUMPTION				CO2 EMISSIO
YEAR		# = high output engine		(L)			TYPE	CITY (L/100 k	HWY (L/100	COMB (L/10(	COMB (mpg)	(g/km)
2000	ACURA	1.6EL	COMPACT	1.6	4	A4	X	9.2	6.7	8.1	35	186
2000	ACURA	1.6EL	COMPACT	1.6	4	M5	X	8.5	6.5	7.6	37	175
2000	ACURA	3.2TL	MID-SIZE	3.2	6	AS5	Z	12.2	7.4	10	28	230
2000	ACURA	3.5RL	MID-SIZE	3.5	6	A4	Z	13.4	9.2	11.5	25	264
2000	ACURA	INTEGRA	SUBCOMPAC	1.8	4	A4	X	10	7	8.6	33	198
2000	ACURA	INTEGRA	SUBCOMPAC	1.8	4	M5	X	9.3	6.8	8.2	34	189
2000	ACURA	INTEGRA GSI	SUBCOMPAC	1.8	4	M5	Z	9.4	7	8.3	34	191

- Fewer errors/omissions/outliers, most attributes
- No location specific variation available
- Already tabulated (CSV)
- 2-cycle & 5-cycle adjusted datasets



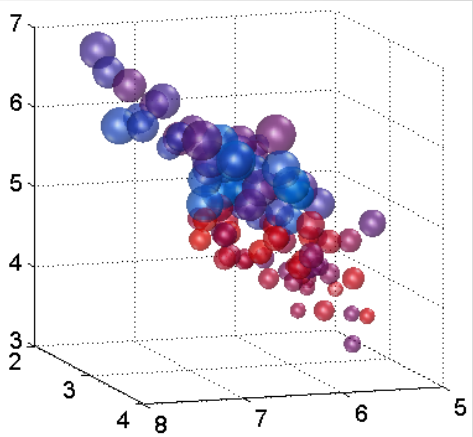
		MODEL(# = high output	VEHICLE	ENGINE		TRANSMISS		FUEL CONSUMPTION				CO2 EMISSIONS
MODELYE	MAKE	engine)	CLASS	SIZE (L)	CYLINDEF	ION	FUEL TYP	CITY (L/100 km)	HWY (L/100 k	COMB (L/100 k	COMB (mp	(g/km)
2014	ACURA	ILX	COMPACT	2	4	AS5	Z	8.6	5.6	7.2	39	166
2014	ACURA	ILX	COMPACT	2.4	4	M6	Z	9.8	6.5	8.3	34	191
2014	ACURA	ILX HYBRID	COMPACT	1.5	4	AV7	Z	5	4.8	4.9	58	113
2014	ACURA	MDX 4WD	SUV - SMALL	3.5	6	AS6	Z	11.2	7.7	9.6	29	221
2014	ACURA	RDX AWD	SUV - SMALL	3.5	6	AS6	Z	10.7	7.3	9.2	31	212
2014	ACURA	RLX	MID-SIZE	3.5	6	AS6	Z	10.5	6.4	8.6	33	198
2014	ACURA	TL	MID-SIZE	3.5	6	AS6	Z	10.4	6.8	8.8	32	202

Analytical Questions

1. Does fuel economy continue to increase year over year for all makes and models?
2. Does fuel economy increased until some common year (1996-98 catalytic converter?), and show signs of a plateau since?
3. Does fuel economy differ significantly by location?
    a. globally, (cheap gas prices in some countries?)
    b. within a country (hilly vs. flat areas)
4. Does real world mean fuel economy match 2-cycle test or 5-cycle test institutional data?

Methods

- Excel & manual examination
- Python pandas, re, urllib2, pool
- Mean of interquartile range for real world data
- Time series analysis
- 2D & 3D visualizations
- Tableau world plots

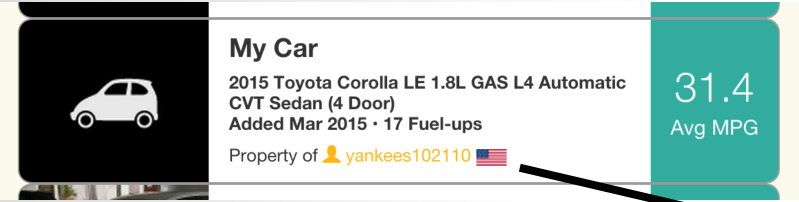


About Tableau maps: www.tableausoftware.com/mapdata
Ecomod Modeling School Alumni

Challenges

- Institutional data (open.canada.ca) not always representative of real world use. (Veracity)
- Real world (fuelly.com) data full of errors, outliers, and missing data. (Veracity, Variety, ~Volume)
- Dimensionality, obtaining enough different make-models, for enough years, with large enough N, with enough fuel ups to provide significant sample.
- Still learning pandas and error-free web scrapping.
- Statistical analysis easy to make assumptions

Real World Data Challenges

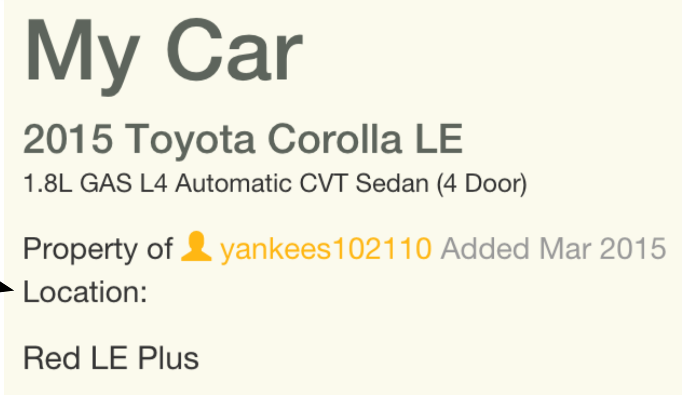


My Car

2015 Toyota Corolla LE 1.8L GAS L4 Automatic CVT Sedan (4 Door)
Added Mar 2015 · 17 Fuel-ups
Property of yankees102110

31.4
Avg MPG

- Location data not always made public, or inaccurate.
- Cleanup, errors, outliers
- Driver predisposition

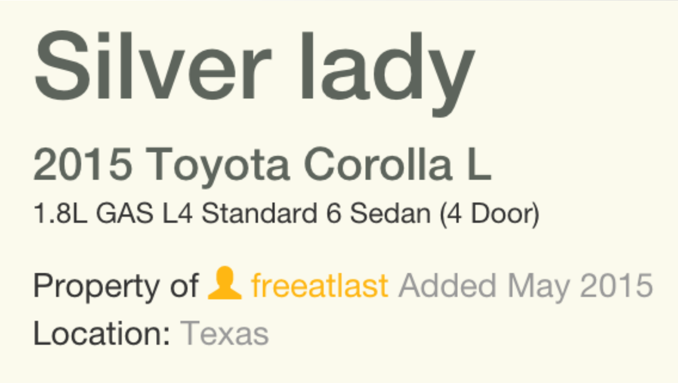


My Car

2015 Toyota Corolla LE
1.8L GAS L4 Automatic CVT Sedan (4 Door)

Property of yankees102110 Added Mar 2015
Location:

Red LE Plus



Silver lady

2015 Toyota Corolla L
1.8L GAS L4 Standard 6 Sedan (4 Door)

Property of freeatlast Added May 2015
Location: Texas

Institutional Data Challenges

“collected from manufacturers who use a specific government approved two-cycle test methodology”

“The new test methods… integrate three additional test cycles that account for air conditioner use, cold temperature operation and driving at higher speeds with more rapid acceleration and braking. In most cases… the new test ratings are 10 to 20 percent higher than the old ratings because they … better approximate everyday driving.”

Timeline

June 3rd - June 26th

- Sampling data sources, examining data attributes for evaluation
- Forming questions
- Determining methods,
- Establish realistic milestones and scopes

Timeline

June 27th - July 17th

- Use institutional data (open.canada.ca) for developing time series analysis, & visualizations (Q1 & Q2)
- Build web scraper for real world data (fuelly.com), generate full tabulated dataset (csv), clean data & summarize in comparable format (Q1 & Q2)
- Generate location plot and analyse data for patterns (Q3)
- Compare institutional (open.canada.ca) data vs real world data (fuelly.com) (Q4)

References

Motorcycle Fuel Economy vs Engine Size

http://web.cs.dal.ca/~dneil/fuelly.php

Institutional Data

http://open.canada.ca/data/en/dataset/98f1a129-f628-4ce4-b24d-6f16bf24dd64

Real World Data

http://www.fuelly.com/car/toyota/corolla/2015

World Plot http://cs.smith.edu/dftwiki/index.php/Geo-Mapping_Data_using_Tableau

3D Scatter Sphere Plot

http://stackoverflow.com/questions/25435174/how-to-visualize-multiple-spheres-with-arrays-of-there-co-ordinate-position-and