CDMA 3654
GROUP 2
Joe Acanfora, Jeff Morris, Sean Cordrey, Jae Choi
Data Manual

Provide a short data manual with the explanations: where the data came from; what is the data set about; what are the variables; what is the meaning of rows/observations (i.e people, cars etc).

Fuel Economy Data

http://catalog.data.gov/dataset/fuel-economy-data

vehicles.csv has 35409 rows and 74 columns.

“Fuel economy data are the result of vehicle testing done at the Environmental Protection Agency's National Vehicle and Fuel Emissions Laboratory in Ann Arbor, Michigan, and by vehicle manufacturers with oversight by EPA. The Find a Car vehicle table contains fuel economy information for 1984-current model year vehicles. The data are available for download in CSV and XML formats:”

The variables names (headers)

- atvtype - type of alternative fuel or advanced technology vehicle
- barrels08 - annual petroleum consumption in barrels for fuelType1 (1)
- barrelsA08 - annual petroleum consumption in barrels for fuelType2 (1)
- charge120 - time to charge an electric vehicle in hours at 120 V
- charge240 - time to charge an electric vehicle in hours at 240 V
- city08 - city MPG for fuelType1 (2)
- city08U - unrounded city MPG for fuelType1 (2), (3)
- cityA08 - city MPG for fuelType2 (2)
- cityA08U - unrounded city MPG for fuelType2 (2), (3)
- cityCD - city gasoline consumption (gallons/100 miles) in charge depleting mode (4)
- cityE - city electricity consumption in kw-hrs/100 miles
- cityUF - EPA city utility factor (share of electricity) for PHEV
- co2 - tailpipe CO2 in grams/mile for fuelType1 (5)
- co2A - tailpipe CO2 in grams/mile for fuelType2 (5)
- co2TailpipeAGpm - tailpipe CO2 in grams/mile for fuelType2 (5)
- co2TailpipeGpm- tailpipe CO2 in grams/mile for fuelType1 (5)
- comb08 - combined MPG for fuelType1 (2)

- comb08U - unrounded combined MPG for fuelType1 (2), (3)
- combA08 - combined MPG for fuelType2 (2)
- combA08U - unrounded combined MPG for fuelType2 (2), (3)
- combE - combined electricity consumption in kw-hrs/100 miles
- combinedCD - combined gasoline consumption (gallons/100 miles) in charge depleting mode (4)
- combinedUF - EPA combined utility factor (share of electricity) for PHEV
- cylinders - engine cylinders
- displ - engine displacement in liters
- drive - drive axle type
- emissionsList
- engId - EPA model type index
- eng_dscr - engine descriptor; see http://www.fueleconomy.gov/feg/findacarhelp.shtml#engine
- evMotor - electric motor (kw-hrs)
- feScore - EPA Fuel Economy Score (-1 = Not available)
- fuelCost08 - annual fuel cost for fuelType1 ($) (7)
- fuelCostA08 - annual fuel cost for fuelType2 ($) (7)
- fuelType - fuel type with fuelType1 and fuelType2 (if applicable)
- fuelType1 - fuel type 1. For single fuel vehicles, this will be the only fuel. For dual fuel vehicles, this will be the conventional fuel.
- fuelType2 - fuel type 2. For dual fuel vehicles, this will be the alternative fuel (e.g. E85, Electricity, CNG, LPG). For single fuel vehicles, this field is not used
- ghgScore - EPA GHG score (-1 = Not available)
- ghgScoreA - EPA GHG score for dual fuel vehicle running on the alternative fuel (-1 = Not available)
- guzzler- if G or T, this vehicle is subject to the gas guzzler tax
- highway08 - highway MPG for fuelType1 (2)
- highway08U - unrounded highway MPG for fuelType1 (2), (3)
- highwayA08 - highway MPG for fuelType2 (2)
- highwayA08U - unrounded highway MPG for fuelType2 (2),(3)
- highwayCD - highway gasoline consumption (gallons/100miles) in charge depleting mode (4)
- highwayE - highway electricity consumption in kw-hrs/100 miles
- highwayUF - EPA highway utility factor (share of electricity) for PHEV
- hlv - hatchback luggage volume (cubic feet) (8)
- hpv - hatchback passenger volume (cubic feet) (8)
- id - vehicle record id
- lv2 - 2 door luggage volume (cubic feet) (8)
- lv4 - 4 door luggage volume (cubic feet) (8)
- make - manufacturer (division)
- mfrCode - 3-character manufacturer code

- model - model name (carline)
- mpgData - has My MPG data; see yourMpgVehicle and yourMpgDriverVehicle
- phevBlended - if true, this vehicle operates on a blend of gasoline and electricity in charge depleting mode
- pv2 - 2-door passenger volume (cubic feet) (8)
- pv4 - 4-door passenger volume (cubic feet) (8)
- rangeA - EPA range for fuelType2
- rangeCityA - EPA city range for fuelType2
- rangeHwyA - EPA highway range for fuelType2
- trans_dscr - transmission descriptor; see http://www.fueleconomy.gov/feg/findacarhelp.shtml#trany
- trany - transmission
- UCity - unadjusted city MPG for fuelType1; see the description of the EPA test procedures
- UCityA - unadjusted city MPG for fuelType2; see the description of the EPA test procedures
- UHighway - unadjusted highway MPG for fuelType1; see the description of the EPA test procedures
- UHighwayA - unadjusted highway MPG for fuelType2; see the description of the EPA test procedures
- VClass - EPA vehicle size class
- year - model year
- youSaveSpend - you save/spend over 5 years compared to an average car ($). Savings are positive; a greater amount spent yields a negative number. For dual fuel vehicles, this is the cost savings for gasoline.
- sCharger - if S, this vehicle is supercharged
- tCharger - if T, this vehicle is turbocharged

emissions

- emissionsList
    - emissionsInfo -
        - efid - engine family ID
        - id - vehicle record ID (links emission data to the vehicle record)
        - salesArea - EPA sales area code
        - score - EPA 1-10 smog rating for fuelType1
        - scoreAlt - EPA 1-10 smog rating for fuelType2
        - smartwayScore - SmartWay Code
        - standard - Vehicle Emission Standard Code
        - stdText - Vehicle Emission Standard

fuel prices

- fuelPrices
    - midgrade - $ per gallon of midgrade gasoline(9)
    - premium - $ per gallon of premium gasoline(9)
    - regular - $ per gallon of regular gasoline(9)
    - cng - $ per gallon of gasoline equivalent (GGE) of compressed natural gas(10)
    - diesel - $ per gallon of diesel(9)
    - e85 - $ per gallon of E85(10)
    - electric - $ per kw-hr of electricity(10)
    - lpg - $ per gallon of propane(10)

yourMpgVehicle - summary of all My MPG data for this vehicle

- avgMpg - harmonic mean of average MPG shared by fueleconomy.gov users
- cityPercent - average % city miles
- highwayPercent - average % highway miles
- maxMpg - maximum user average MPG
- minMpg - minimum user average MPG
- recordCount - number of records for this vehicle
- vehicleId - vehicle record id (links My MPG data to the vehicle record)

yourMpgDriverVehicle - summary of driver data reported for this vehicle

- cityPercent - user average % city miles
- highwayPercent - user average % highway miles
- lastDate - date records were last updated (yyyy-mm-dd)
- mpg - average MPG
- state - state of residence
- vehicleId - vehicle record ID (links My MPG data to the vehicle record)

Statistical Summaries

```
#numeric variables
# 0 barrels08
summary(vehicleData$barrels08)
   #    Min.  1st Qu.   Median     Mean  3rd Qu.     Max.
   #0.05989 14.96000 17.34000 17.72000 20.59000 47.07000
# 1 barrelsA08
summary(vehicleData$barrelsA08)
```

```
   #   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
   # 0.0000  0.0000  0.0000  0.2081  0.0000 18.3000
# 2 charge120
summary(vehicleData$charge120)
   #Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
   #0       0       0       0       0       0
# 3 charge240
summary(vehicleData$charge240)
   #Min.  1st Qu.   Median     Mean  3rd Qu.     Max.
   #0.00000  0.00000  0.00000  0.01035  0.00000 12.00000
# 4 city08
summary(vehicleData$city08)
   #  Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
   #6.00   15.00   17.00   17.64   20.00  138.00
# 5 city08U
summary(vehicleData$city08U)
   #    Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
   #0.000   0.000   0.000   2.788   0.000 138.300
# 6 cityA08
summary(vehicleData$cityA08)
  #    Min.  1st Qu.   Median     Mean  3rd Qu.     Max.
  #0.0000   0.0000   0.0000   0.4389   0.0000 127.0000
# 7 cityA08U
summary(vehicleData$cityA08U)
   #    Min.  1st Qu.   Median     Mean  3rd Qu.     Max.
   #0.0000   0.0000   0.0000   0.2321   0.0000 127.1000
# 8 cityCD
summary(vehicleData$cityCD)
   #    Min.  1st Qu.   Median     Mean  3rd Qu.     Max.
   #0.000000 0.000000 0.000000 0.000391 0.000000 5.350000
# 9 cityE
summary(vehicleData$cityE)
   #    Min.  1st Qu.   Median     Mean  3rd Qu.     Max.
   #0.0000   0.0000   0.0000   0.1085   0.0000 122.0000
# 10 cityUF
summary(vehicleData$cityUF)
   #     Min.   1st Qu.    Median      Mean   3rd Qu.      Max.
   #0.0000000 0.0000000 0.0000000 0.0003527 0.0000000 0.8490000
# 11 co2
summary(vehicleData$co2)
   #Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
   #-1.00   -1.00   -1.00   38.59   -1.00  847.00
# 12 co2A
summary(vehicleData$co2A)
   #   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
   #-1.0    -1.0    -1.0     4.1    -1.0   719.0
```

```
# 13 co2TailpipeAGpm
summary(vehicleData$co2TailpipeAGpm)
   #   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
   #0.00    0.00    0.00   17.31    0.00  719.00
# 14 co2TailpipeGpm
summary(vehicleData$co2TailpipeGpm)
   #Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
   #0.0   404.0   467.7   478.6   555.4  1270.0
# 15 comb08
summary(vehicleData$comb08)
   #   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
   #7.00   16.00   19.00   19.87   22.00  124.00
summary(vehicleData$comb08U)
#Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
#0.000   0.000   0.000   3.151   0.000 124.400
summary(vehicleData$combA08)
#Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
#0.000   0.000   0.000   0.497   0.000 117.000
summary(vehicleData$combA08U)
#Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
#0.000   0.000   0.000   0.497   0.000 117.000
summary(vehicleData$combE)
#Min.  1st Qu.   Median     Mean  3rd Qu.     Max.
#0.0000   0.0000   0.0000   0.1138   0.0000 121.0000
summary(vehicleData$combinedCD)
#Min.  1st Qu.   Median     Mean  3rd Qu.     Max.
#0.000000 0.000000 0.000000 0.000324 0.000000 4.800000
summary(vehicleData$combinedUF)
#Min.   1st Qu.    Median     Mean   3rd Qu.     Max.
#0.0000000 0.0000000 0.0000000 0.0003412 0.0000000 0.8340000
summary(vehicleData$cylinders)
#Min. 1st Qu.  Median    Mean 3rd Qu.    Max.    NA's
#  2.000   4.000   6.000   5.744   6.000  16.000      71
summary(vehicleData$drive)
#2-Wheel Drive           4-Wheel Drive
#1189                  507                  768
#4-Wheel or All-Wheel Drive          All-Wheel Drive        Front-Wheel Drive
#6648                 1444                 12504
#Part-time 4-Wheel Drive         Rear-Wheel Drive
#112                12236
summary(vehicleData$engId)
#Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
#0      0    309    9455    4820   69100
summary(vehicleData$eng_dscr)
#(FFS)
#14690                  8827
```

#SIDI (FFS) CA model
#2089 926
#(FFS) (MPFI) (FFS,TRBO)
#734 666
#FFV (350 V8) (FFS)
#572 411
#(GUZZLER) (FFS) SOHC
#366 354
#(NO-CAT) FLEX-FUEL
#238 198
#GUZZLER (FFS) (SPFI)
#195 194
#SIDI; FFV (GUZZLER) (FFS) (MPFI)
#181 122
#(350 V8) CA model
#120 113
#(350 V8) (FFS) (MPFI) (GM-CHEV)
#106 102
#DOHC (FFS) (DIESEL)
#96 95
#PR (GUZZLER) (FFS)
#91 84
#(FFS,TRBO) CA model DOHC
#81 79
#SOHC (FFS) DOHC TURBO (FFS,TRBO)
#78 76
#V-6 (305) (FFS)
#75 71
#(DIESEL) CA model (CAL)(FFS)
#71 67
#(DSL,TRBO) SOHC-4 (FFS)
#60 50
#HEV (GM-CHEV) (FFS)
#49 46
#(CALIF) (DOHC) (FFS)
#45 44
#(GUZZLER) DOHC-IL4
#42 42
#(305) (GUZZLER) (FFS,TRBO)
#40 40
#(SOHC) (FFS) LM7
#39 38
#(DSL,TRBO) (NO-CAT) (FFS) fuel injection
#37 37
#SOHC-4 2WD (FFS) FFS
#37 34

#VTEC (4A-FE) (FFS)
#33 32
#(FFS) (S-CHARGE) SOHC-4 4WD (FFS)
#32 32
#(350 V8) (DIESEL) (350 V8) (FFS) CA model
#31 30
#(FFS) 2 barrel carb (FFS,TRBO) (MPFI)
#29 29
#(POLICE) (FFS) SIDI & PFI
#28 28
#V6 (307) (FFS)
#28 27
#SOHC-VTEC (MPFI) (NO-CAT)
#27 26
#B235R DOHC-VTEC
#26 25
#(DSL,TRBO) (MPFI) (DSL,TRBO) CA model
#24 24
#4V (GM-CHEV) CA model
#24 23
#PHEV (3S-FE) (FFS)
#23 21
#(GM-OLDS) (FFS) (DIESEL) (NO-CAT)
#21 20
#(FFS) (GUZZLER) (FFS) (MPFI)
#20 20
#GAS 330 I4
#20 20
#4-VALVE (16-VALVE) (FFS)
#19 18
#MOTORSPORT SPORTS
#18 18
#(CALIF) (FFS,TRBO) (VTEC) (FFS)
#17 17
#L410MT2 SIDI; with Stop-Start Option
#17 17
#(121) (FFS) (GUZZLER) (FFS,TRBO) (MPFI)
#16 16
#(GUZZLER) CA model (MPFI)
#16 16
#4 VALVE B235E
#16 16
#DOHC-T/C L-4
#16 16
#(DOHC) (FFS,TRBO) (GM-CHEV) (FFS) CA model
#15 15

```
#(122)      (FFS)           (16VALVES) (FFS)
#14                      14
#2-VALVE                   390-540
#14                      14
#B205R                    (Other)
#14                     1468
summary(vehicleData$feScore)
#Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
#-1.0000 -1.0000 -1.0000 -0.3924 -1.0000 10.0000
summary(vehicleData$fuelCost08)
#Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
#500    2350    2750    2864    3250    8150
summary(vehicleData$fuelCostA08)
#Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
#0.0     0.0     0.0   141.4     0.0  5700.0
summary(vehicleData$fuelType)
#CNG                    Diesel
#59                     1054
#Electricity            Gasoline or E85
#68                     1088
#Gasoline or natural gas        Gasoline or propane
#18                       8
#Midgrade                   Premium
#48                     8923
#Premium and Electricity  Premium Gas or Electricity
#5                      10
#Premium or E85                 Regular
#99                    24018
#Regular Gas and Electricity
#10
summary(vehicleData$fuelType1)
#Diesel       Electricity Midgrade Gasoline       Natural Gas
#1054              68              48              59
#Premium Gasoline  Regular Gasoline
#9037           25142
summary(vehicleData$ghgScore)
#Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
#-1.0000 -1.0000 -1.0000 -0.3941 -1.0000 10.0000
summary(vehicleData$ghgScoreA)
#Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
#-1.0000 -1.0000 -1.0000 -0.9471 -1.0000  8.0000
summary(vehicles$highway08)
#Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
#9.00   20.00   23.00   23.74   27.00  111.00
summary(vehicles$highway08U)
#Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
```

```
#0.000   0.000   0.000   3.788   0.000 111.400
summary(vehicles$highwayA08)
#Min.  1st Qu.   Median     Mean  3rd Qu.     Max.
#0.0000   0.0000   0.0000   0.5963   0.0000 107.0000
summary(vehicles$highwayA08U)
#Min.  1st Qu.   Median     Mean  3rd Qu.     Max.
#0.0000   0.0000   0.0000   0.3092   0.0000 106.5000
summary(vehicles$highwayCD)
#Min.  1st Qu.   Median     Mean  3rd Qu.     Max.
#0.000000 0.000000 0.000000 0.000235 0.000000 4.060000
summary(vehicles$highwayE)
#Min.  1st Qu.   Median     Mean  3rd Qu.     Max.
#0.0000   0.0000   0.0000   0.1207   0.0000 120.0000
summary(vehicles$highwayUF)
#Min.   1st Qu.    Median      Mean   3rd Qu.      Max.
#0.0000000 0.0000000 0.0000000 0.0003275 0.0000000 0.8130000
summary(vehicles$hlv)
# Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
#0.00    0.00    0.00    2.04    0.00   49.00
summary(vehicles$hpv)
#Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
#0.0     0.0     0.0    10.5     0.0   195.0
summary(vehicles$id)
#Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
#1    8853   17710   17790   26740   35710
summary(vehicles$lv2)
#Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
#0.000   0.000   0.000   1.875   0.000  41.000
summary(vehicles$lv4)
#Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
#0.000   0.000   0.000   6.217  13.000  55.000
summary(vehicles$pv2)
#Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
#0.00    0.00    0.00   13.75    0.00  194.00
summary(vehicles$pv4)
#Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
#0.00    0.00    0.00   13.75    0.00  194.00
summary(vehicles$range)
#Min.  1st Qu.   Median     Mean  3rd Qu.     Max.
#0.0000   0.0000   0.0000   0.1705   0.0000 265.0000
summary(vehicles$rangeCity)
#Min.  1st Qu.   Median     Mean  3rd Qu.     Max.
#0.0000   0.0000   0.0000   0.1246   0.0000 262.7000
summary(vehicles$rangeCityA)
#Min.  1st Qu.   Median     Mean  3rd Qu.     Max.
#0.00000  0.00000  0.00000  0.01829  0.00000 77.50000
```

```
summary(vehicles$rangeHwy)
#Min.  1st Qu.   Median     Mean  3rd Qu.     Max.
#0.0000   0.0000   0.0000   0.1152   0.0000 266.8000
summary(vehicles$UCity)
#Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
#0.00   18.00   21.00   22.17   25.00  197.60
summary(vehicles$UCityA)
#Min.  1st Qu.   Median     Mean  3rd Qu.     Max.
#0.0000   0.0000   0.0000   0.5393   0.0000 181.6000
summary(vehicles$UHighway)
#Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
#0.00   27.00   32.45   33.08   37.76  159.10
summary(vehicles$UHighwayA)
#Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
#0.000   0.000   0.000   0.803   0.000 152.200
summary(vehicles$year)
#Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
#1984    1990    1999    1999    2008    2015
summary(vehicles$youSaveSpend)
#Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
#-29500   -5000   -2500   -3071    -500    8750


# non-numeric
# 1 make
# 2 model
# 3 mpgData
# 4 phevBlend
# 5 trany
# 6 guzzler - incomplete the blank data should be F or false
# 7 trans_scr - incomplete
# 8 tCharger - incomplete the blank data should be F or false
# 9 sCharger - incomplete the blank data should be F or false
# 10 atvType - incomplete the blank data should be "gas" or "standard"
# 11 fuelType2 - incomplete
# 12 rangeA - incomplete
# 13 evMotor - incomplete
# 14 mfrCode - incomplete
```