

R Lecture 1

設定工作目錄

```
getwd()  
setwd("C:\\Program Files\\R\\working")
```

讀取資料

`read.table()` .txt, 以空白(" ")或 Tab("")做分隔。
`read.csv()` .csv, 以 , 分隔。

匯出資料

```
write.csv(df, "df.csv", sep=",", col.names=TRUE)  
write.table(df, "df.txt", quote=FALSE, sep="\t")
```

預覽資料

`head()`, 前 6 筆
`tail()`, 後 6 筆

summarize 資料

`str()` 列出資料內每個欄位的狀態
`summary()` 列出 5-number summary & mean
`names()`, `colnames()` 得到變數名稱
`class()` 得到型態
`table()`, `mean()`, `min()`, `max()`, `range()`, `var()`, `sd()`, `cov()` `quantile()`,
`quantile(變數, c(0.3, 0.7, 0.99))` `dim()` 給幾個 row 幾個 column

套件

`install.packages()` 裝套件

`library()` 匯入套件

向量化運算

`lapply(資料, fnc), sapply()`

畫圖

- `plot(V1, V2)`

- `library(lattice)`

`xyplot(V2 ~ V1, data)` 以公式的方式來指定 x 與 y 軸的資料

`xyplot(V2 ~ V1 | 類別變數, data)`

- `library(ggplot2)`

`ggplot(data, aes(x = V1, y = V2)) +`

`geom_point()`

<http://www.sthda.com/english/wiki/data-visualization>

缺值

`is.na()` 是否有缺值

`data[!complete.cases(data),]` list rows of data that have missing values

`newdata <- na.omit(data)` create new dataset without missing data

`na.omit()` omit all rows containing missing values.

```
data <- c(12, 4, 6, NA, 45, NA, 99)
```

```
is.na(data) #是否是NA
```

```
## [1] FALSE FALSE FALSE TRUE FALSE TRUE FALSE
```

```
which(is.na(data)) #哪個是NA
```

```
## [1] 4 6
```

```
sum(is.na(data)) #幾個NA
```

```
## [1] 2

mean(data) #有NA 無法跑

## [1] NA

mean(data, na.rm = TRUE) #除掉NA 算

## [1] 33.2

data[is.na(data)] <- mean(data, na.rm = TRUE) #NA 用mean 取代
data

## [1] 12.0  4.0  6.0 33.2 45.0 33.2 99.0

data[data == 99] <- NA #99 改成NA
data

## [1] 12.0  4.0  6.0 33.2 45.0 33.2  NA
```

例子-Car Sale

```
setwd("~/Downloads")
print(load("carsale.Rda"))

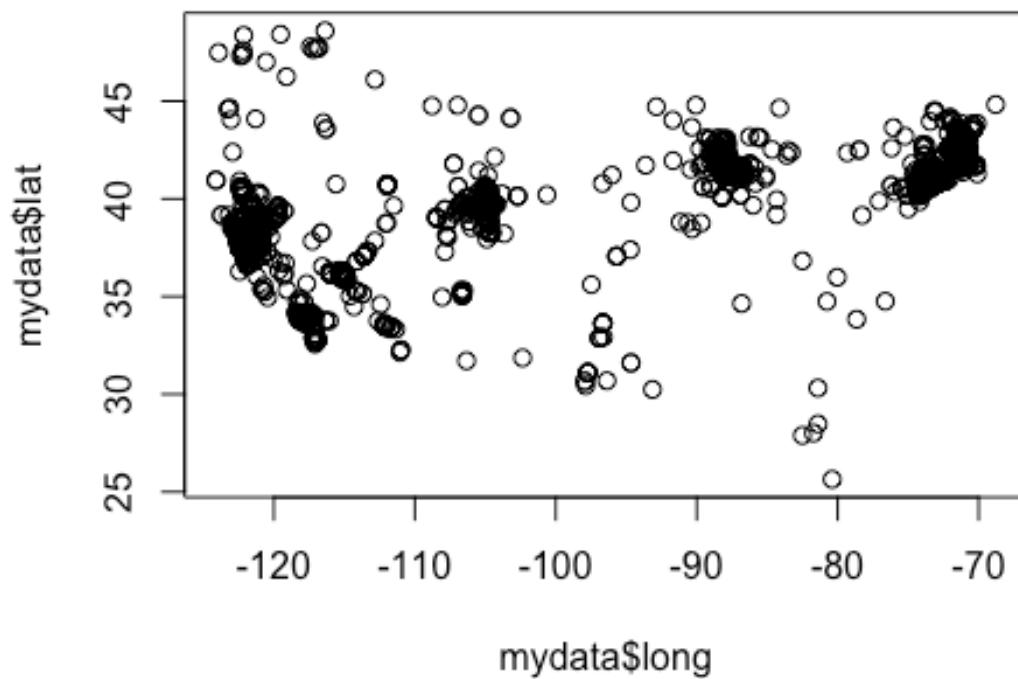
str(mydata)
summary(mydata)
class(mydata)
class(mydata[,1])
sapply(mydata, class)

table(mydata$maker)

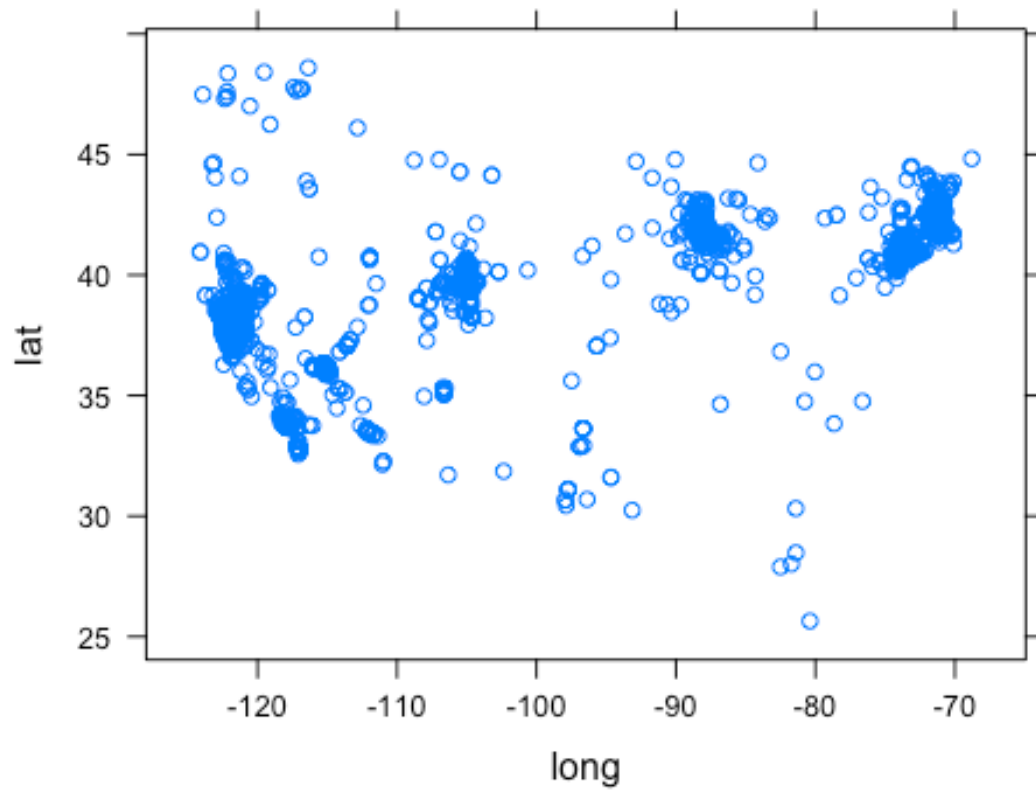
##
##          acura          alfa romeo          amc          aston martin
##          697             8             6             3
##          audi          bentley          bmw          bricklin
##          579             15          1657             1
##          bugatti          buick          cadillac          chevrolet
##          2             363             571          3394
##          chrysler          daewoo          datsun          desoto
##          835             5             16             2
##          dodge          eagle          ferrari          fiat
##          1841             6             4             35
##          ford          freightliner          geo          gmc
##          4266             39             24          684
## harley davidson          honda          hudson          hummer
##          15             2650             6             50
##          hyundai          infiniti          international          isuzu
##          876             559             66             50
##          jaguar          jeep          kia          lamborghini
```

```
##          127          1022          395          3
##    land rover      leaf      lexus    lincoln
##          148           8       786       307
##      mack      maserati      mazda    mercedes
##          7          17       550       1283
##    mercury      mg      mini    mitsubishi
##        230          19       215       363
##      nissan    oldsmobile    peterbilt    peugeot
##        2473          108           9           1
##    plymouth    pontiac    porsche    rolls royce
##          58          431          162          10
##      saab      saturn      scion      shelby
##        114          220          151          13
##      smart    studebaker    subaru      suzuki
##          25           5       531          93
##      tesla      toyota      triumph    volkswagen
##          9          3332           7       1116
##      volvo      willys    yerfdog      zap
##        371          12           2           1
```

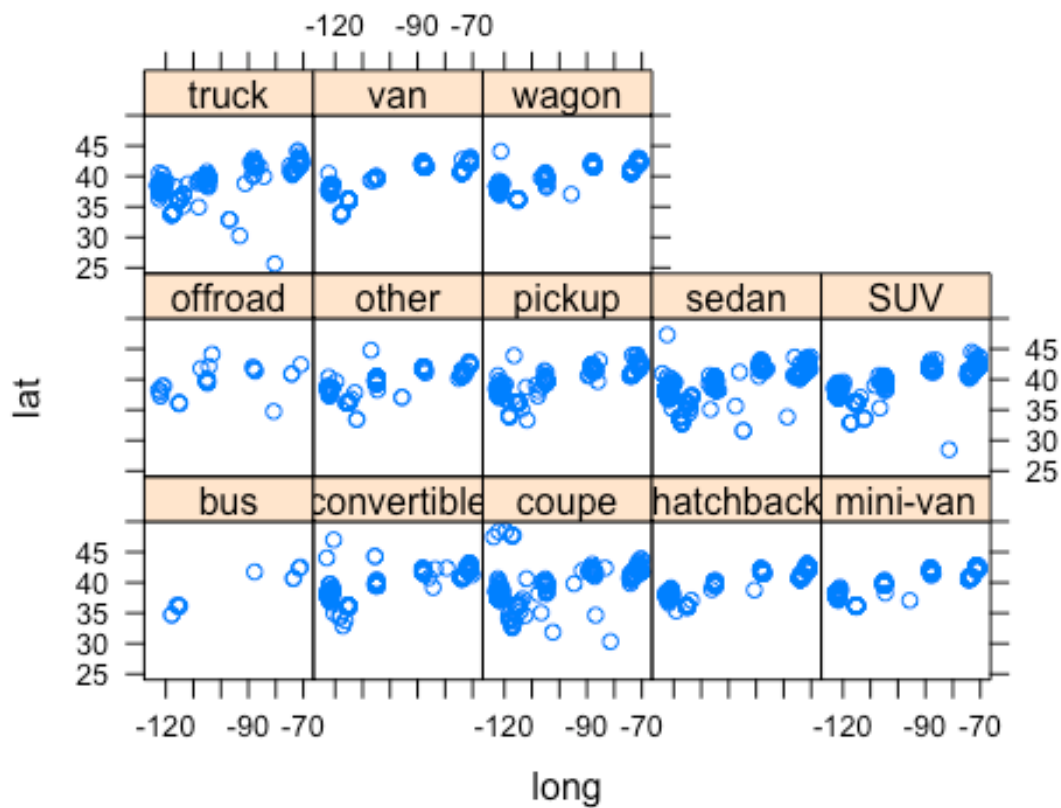
```
plot(mydata$long,mydata$lat)
```



```
library(lattice)
xyplot(lat ~ long, mydata)
```

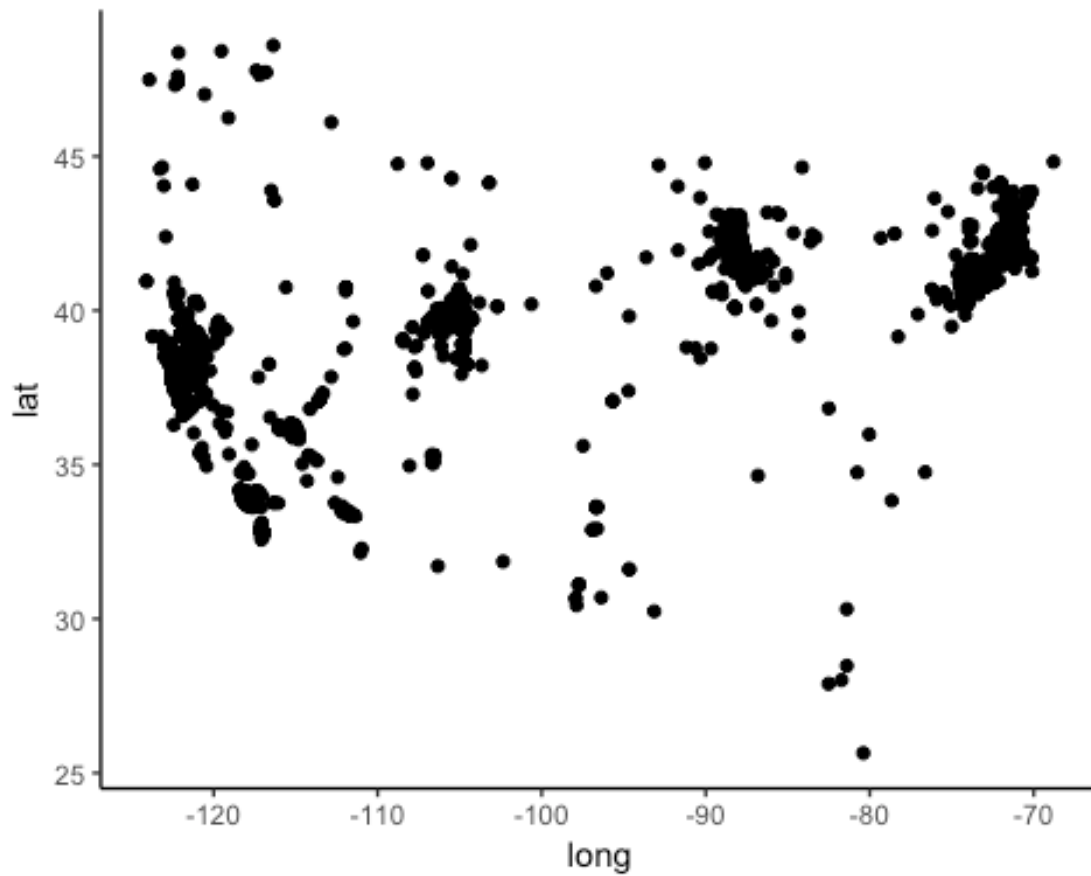


```
xyplot(lat ~ long | type, mydata)
```



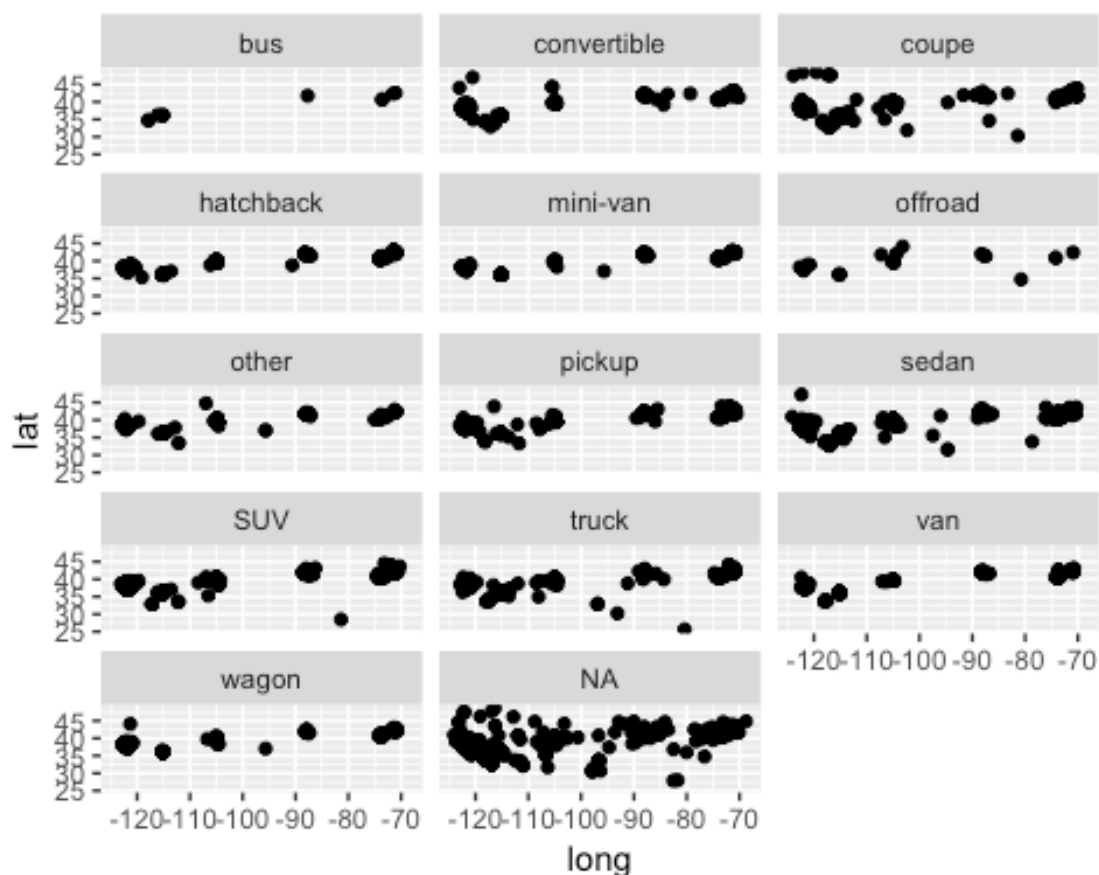
```
#install.packages("ggplot2")
library(ggplot2)
ggplot(mydata, aes(x = long, y = lat)) +
  geom_point()+
  theme_classic()

## Warning: Removed 14445 rows containing missing values (geom_point).
```



```
ggplot(mydata, aes(x = long, y = lat)) +  
  geom_point() +  
  facet_wrap(~ type, ncol = 3)
```

```
## Warning: Removed 14445 rows containing missing values (geom_point).
```



```
df_type = mydata[which(mydata$type != "NA"),] #去掉type 有NA 的
df = na.omit(mydata) #去掉所有NA
summary(df$city)

##    boston    chicago    denver lasvegas      nyc      sac      sfbay
##      459       290       336       312      346      287      362

table(df$city)

##
##    boston    chicago    denver lasvegas      nyc      sac      sfbay
##      459       290       336       312      346      287      362

levels(df$city)

## [1] "boston"  "chicago" "denver"   "lasvegas" "nyc"      "sac"
## [7] "sfbay"
```

最貴的車？

```
max(mydata$price,na.rm=T)

## [1] 30002500
```



```

mydata[which.max(mydata$price),]$body

## [1] "\n          clean, fully loaded, nice shine, good running engine
and trans, willing to trade for old school or truck??????????????? M
ounted on 22 inch rims new tires no bends no cracks\n          "

median(c(6000,30000))

## [1] 18000

mydata[which(mydata$price==max(mydata$price,na.rm=T)),]$price=median(c
(6000,30000)) #用中位數取代
mydata[which.max(mydata$price),]$header #哪部車最貴

## [1] "2001 Honda Accord"

head(sort(mydata$price,decreasing=T))

## [1] 9999999  569500  559500  400000  359000  286763

```

子集合

```

bmw_df <- subset(df, maker=="bmw")
head(bmw_df)

##              id
## posted23  5229007234
## posted100 5230459819
## posted552 5229211998
## posted554 5229216463
## posted817 5230493753
## posted934 5231714466
##
##              title
## posted23  2008 BMW X5 3.0si,,ONLY 71K,,LIKE NEW,,1OWNER,SPORT PKG,NA
VI,NEW TIRES - $19997 (Ashland)
## posted100 2008 BMW X5 3.0si,,ONLY 71K,,LIKE NEW,,1OWNER,SPORT PKG,NA
VI,NEW TIRES - $19997 (Ashland)
## posted552          1999 BMW Z3M COUPE, SUPER CLEAN, ADULT OWNED, JUS
T SERVICED - $19500 (ATTLEBORO)
## posted554          2002 BMW 330XI, SUPER CLEAN, ALL WHEEL DRI
VE, M SPORT - $6999 (ATTLEBORO)
## posted817          2007 BMW 525Xi - $9975
(Legendary Motors   Braintree)
## posted934          2013 BMW X5 PREMIUM SPRT.22K MILES ONE OWNER CLEAN CA
RFAX.LIKE NEW - $38990 (DEDHAM)
##

```

body

posted23

\n 1OWNER 2008 X5 3.0SI IN A MINT CONDITION
(HARD TO FIND WHITE ON BEIGE LEATHER). CLEAN CARFAX REPORT WITH ALL MA
INTENANCE RECORDS.ALL SERVICES WAS DONE @BMW DEALER. THIS CAR HAS NO AC
CIDENTS,NO PAINTWORK.\nNEW TIRES JUST REPLACED!!! DRIVES AND LOOKS LIKE
NEW CAR!!!\nTHIS IS NON-SMOKING VEHICLE,NO DINGS,NO DENTS NO SCRATCHES.
\nTHIS X5 IS LOADED:\nSPORT PKG.\nADAPTIVE DRIVE.\nNAVIGATION.\nCOLD WE
ATHER PKG(HEATED FRONT SEATS AND STEERING WHEEL)\nPANORAMIC ROOF.\nREAR
CLIMATE CONTROL.\nPARKING SENSORS.\nBI-XENON HEADLIGHTS.\nHI-FI SOUND
SYSTEM.\nSPORT SEATS.\nREAR SUNSHADES.\nAUTOMATIC FOLDING MIRRORS.\n19'
' ALLOY WHEELS TYPE212\nFACTORY RUNNING BOARDS AND MUCH MORE....\nHARD
TO FIND ANOTHER ONE LIKE THIS!!!\nWITH ANY QUESTIONS CALL OR TEXT ANDRE
Y @617-799-8738\nTHANK YOU.\n\n## posted100

\n 10OWNER 2008 X5 3.0SI IN A MINT CONDITION
(HARD TO FIND WHITE ON BEIGE LEATHER). CLEAN CARFAX REPORT WITH ALL MA
INTENANCE RECORDS.ALL SERVICES WAS DONE @BMW DEALER. THIS CAR HAS NO AC
CIDENTS,NO PAINTWORK.\nNEW TIRES JUST REPLACED!!! DRIVES AND LOOKS LIKE
NEW CAR!!!\nTHIS IS NON-SMOKING VEHICLE,NO DINGS,NO DENTS NO SCRATCHES.
\nTHIS X5 IS LOADED:\nSPORT PKG.\nADAPTIVE DRIVE.\nNAVIGATION.\nCOLD WE
ATHER PKG(HEATED FRONT SEATS AND STEERING WHEEL)\nPANORAMIC ROOF.\nREAR
CLIMATE CONTROL.\nPARKING SENSORS.\nBI-XENON HEADLIGHTS.\nHI-FI SOUND
SYSTEM.\nSPORT SEATS.\nREAR SUNSHADES.\nAUTOMATIC FOLDING MIRRORS.\n19'
' ALLOY WHEELS TYPE212\nFACTORY RUNNING BOARDS AND MUCH MORE....\nHARD
TO FIND ANOTHER ONE LIKE THIS!!!\nWITH ANY QUESTIONS CALL OR TEXT ANDRE
Y @617-799-8738\nTHANK YOU.\n## posted552

\n This is a very nic
e example of a the highly sought after 1999 BMW Z3 M COUPE, 98,500 mile
s, arctic silver exterior, black nappa interior, adult owned/ maintaine
d, vehicle has never been raced/abused. This car is very rare in this c
ondition. Total number produced in arctic silver over black nappa for 1
999 with a sunroof was only 105.\nThis M was just serviced by a reputab
le BMW tech and just received a new Bilstein suspension kit, new front
drilled rotors, stainless steel brake lines, rebuilt brake calipers pow
der coated red, fresh oil change, new hatch struts and a few other misc
ellaneous parts. Car has never been in an accident via auto check, clea
n car fax in hand and has zero rust on the exterior, zero tears in the
interior and is very clean overall considering its age/mileage.\nFor an
y other questions email or call 6179811908, thanks.\nWill consider trad
e for air cooled Porsche 911\n## posted554

\n 2002 BMW 330XI, automatic transmission, 115
k miles, adult owned, just serviced by BMW, car runs and drives like ne
w. Trades are accepted and financing is available. For more info please
email or call 6179811908, thanks.

posted817 \n Very nice clean well maintained BMW that is equi
pped nicely options and ready for winter with AWD\nGreat CarFax report;

No accidents and 13 service records available.\nCarFax appraises this
vehicle \$560 more than full Kelley Blue Book retail value of \$10,585\nE
quipped very nicely;\n3.0 Liter V6 Engine\nAutomatic Transmission w/Ste
ptronic\nAll Wheel Drive\nHill Descent Control\nBluetooth Wireless\nDua
l Air Bags\nHID Headlights\nLeather Interior\nAM/FM/CD/MP3\nDual Zone A
ir Conditioning\nDual Power Heated Seats\nPower Glass Moonroof\nPower W
indows\nPower Mirrors\nPower Remote Door Locks\nTilt & Telescopic Wheel
\nCruise Control\nAlloy Wheels\nExcellent Tires and much more!!\nThis i
s what I call a Legendary Motors experience;\n(1) All vehicles are safe
ty checked and serviced on site by certified A.C.E. independent mechani
cs. \n(2) Our advertised price includes a 3 month/3000 mile CornerStone
PowerTrain warranty that is honored nationwide.\nCornerStone is rated
A+ by the Better Business Bureau. A warranty that gives you piece of mi
nd whether going across town or across the country. We also are accredi
ted A+ by the BBB.\n(3) We don't waste your time with the "used car pur
chase show", we have no managers to check with about the best price. \n
For more pictures, warranty information or financing options, please gi
ve us a call anytime. \nFor complete inventory viewing click on "more u
ser ads"\nOffice @ 781-356-0504-- Cellphones Doug @ 857 333 6884 or me,
GK @ 781-589-4103\nWe work with many banks & credit unions to get you
the best rates, in these difficult times people sometimes have damaged
credit so we do work with lenders that offer guaranteed financing.\nLeg
endary Motors\n577 Pond St\nBraintree Ma\n10 am--6pm Monday -- Friday\n10 am--4pm Saturday --Sunday\n

posted934

\n 2013 BMW X5 PREMIUM SUV. 23K MILES. THIS CAR STILL UNDER FACTORY WARRANTY AND LIKE NEW.LOADED premium pkg cold weather pkg NAVIGATION BACKUP CAM.. SUNROOF SIDE STEPS.POWER LIFTGATE. IF YOU HAVE ANY QUETIONS PLEASE CALL PETER AT 781-801-3646..THANKS(no fees or charges). last week at this price.\n

##	lat	long	posted	updated
----	-----	------	--------	---------

drive				
## posted23	42.25173	-71.45323	2015-09-19 21:42:25	2015-09-19 21:42:26
4wd				

## posted100	42.25173	-71.45323	2015-09-20 21:58:50	2015-09-20 21:58:51
4wd				

## posted552	41.92960	-71.30090	2015-09-19 23:50:31	2015-09-19 23:50:32
rwd				

## posted554	41.92960	-71.30090	2015-09-19 23:52:59	2015-09-19 23:53:01
4wd				

## posted817	42.20930	-70.99630	2015-09-20 22:28:03	2015-09-22 09:48:04
4wd				

## posted934	42.24370	-71.16370	2015-09-21 19:47:13	2015-09-21 19:47:14
4wd				

##	odometer	type	header	condition	cylinders	fuel
----	----------	------	--------	-----------	-----------	------

## posted23	71455	SUV	2008 bmw x5 3.0si	like new	6	gas
-------------	-------	-----	-------------------	----------	---	-----

## posted100	71455	SUV	2008 bmw x5 3.0si	like new	6	gas
--------------	-------	-----	-------------------	----------	---	-----

## posted552	98500	coupe	1999 BMW m coupe	excellent	6	gas
--------------	-------	-------	------------------	-----------	---	-----

## posted554	115888	sedan	2002 BMW 330XI	excellent	6	gas
--------------	--------	-------	----------------	-----------	---	-----

```

## posted817 121085 sedan 2007 BMW 525Xi excellent 6
gas
## posted934 22 SUV 2013 bmw x5 xdrive35i excellent 6
gas
## size transmission byOwner city time
## posted23 mid-size automatic FALSE boston 2015-09-20 00:42:00
## posted100 mid-size automatic FALSE boston 2015-09-21 00:58:00
## posted552 compact manual FALSE boston 2015-09-20 02:50:00
## posted554 mid-size automatic FALSE boston 2015-09-20 02:52:00
## posted817 full-size automatic FALSE boston 2015-09-21 01:28:00
## posted934 full-size automatic FALSE boston 2015-09-21 22:47:00
##
description
## posted23 2008 BMW X5 3.0si,,ONLY 71K,,LIKE NEW,,10OWNER,SPORT PKG,NA
VI,NEW TIRES
## posted100 2008 BMW X5 3.0si,,ONLY 71K,,LIKE NEW,,10OWNER,SPORT PKG,NA
VI,NEW TIRES
## posted552 1999 BMW Z3M COUPE, SUPER CLEAN, ADULT OWNED, J
UST SERVICED
## posted554 2002 BMW 330XI, SUPER CLEAN, ALL WHEEL DR
IVE, M SPORT
## posted817 20
07 BMW 525Xi
## posted934 2013 BMW X5 PREMIUM SPRT.22K MILES ONE OWNER CLEAN CAR
FAX.LIKE NEW
## location
## posted23 (Ashland) pic map
## posted100 (Ashland) pic map
## posted552 (ATTLEBORO) pic map
## posted554 (ATTLEBORO) pic map
## posted817 (Legendary Motors Braintree) pic map
## posted934 (DEDHAM) pic map
## url price year maker makerMethod
## posted23 /bmw/ctd/5229007234.html 19997 2008 bmw 1.5
## posted100 /bmw/ctd/5230459819.html 19997 2008 bmw 1.5
## posted552 /gbs/ctd/5229211998.html 19500 1999 bmw 1.5
## posted554 /gbs/ctd/5229216463.html 6999 2002 bmw 1.5
## posted817 /gbs/ctd/5230493753.html 9975 2007 bmw 1.5
## posted934 /gbs/ctd/5231714466.html 38990 2013 bmw 1.5

subset(df, price > 40000, select = c(price, maker,type))

## price maker type
## posted1254 51400 chrysler mini-van
## posted13541 42000 mercedes coupe
## posted14921 49000 dodge coupe
## posted22901 41800 land rover SUV
## posted16515 49000 gmc SUV
## posted21263 45800 chevrolet sedan
## posted3618 43500 lexus SUV

```

```
## posted13245 105000      ford  sedan
## posted17020  43748  chevrolet  truck
## posted18220  41783  chevrolet  truck
## posted7236   45559      ford  truck
## posted7436   43745  chevrolet  truck
## posted7506   41787  chevrolet  truck
## posted5197   41999  chevrolet  coupe
## posted8937   62000      dodge  coupe
## posted16148  48999      gmc     SUV
## posted20339  48000      gmc     SUV
## posted20509  48000  chevrolet  truck
## posted29213  57500  chevrolet  coupe
## posted135812 53444  mercedes  sedan
## posted135912 62444  nissan    coupe
## posted111913 45000      bmw     sedan
```

```
subset(df, price > 40000 & type=='SUV', select = c(price, maker, type))
```

```
##           price      maker type
## posted22901 41800 land rover  SUV
## posted16515 49000      gmc    SUV
## posted3618  43500    lexus    SUV
## posted16148 48999      gmc    SUV
## posted20339 48000      gmc    SUV
```