Initialize your Stata edition

In [2]:
```python
import os
os.chdir("C:/Program Files/Stata17/utilities")
from pystata import config
config.init("se")
```

```
  ___  ____  ____  ____  ____ ©
 /__    /   ____/   /   ____/      17.0
___/   /   /___/   /   /___/       SE—Standard Edition

 Statistics and Data Science       Copyright 1985-2021 StataCorp LLC
                                   StataCorp
                                   4905 Lakeway Drive
                                   College Station, Texas 77845 USA
                                   800-STATA-PC        https://www.stata.com
                                   979-696-4600        stata@stata.com

Stata license: Single-user  perpetual
Serial number: [redacted]
  Licensed to: [redacted]

Notes:
      1. Unicode is supported; see help unicode_advice.
      2. Maximum number of variables is set to 5,000; see help set_maxvar.
```

I began experimenting with the popular Stata system file, the auto.dta dataset is the most popular. It's nice and easy

In [5]:
```
%%stata
sysuse auto, clear
```

```
(1978 automobile data)
```

We begin by inspecting the dataset with the simple describe command.

In [6]:
```
%%stata
describe
```

```
Contains data from C:\Program Files\Stata17/ado\base/a/auto.dta
 Observations:           74                  1978 automobile data
    Variables:           12                  13 Apr 2020 17:45
                                             (_dta has notes)
-------------------------------------------------------------------------------
Variable      Storage   Display    Value
    name         type    format    label      Variable label
-------------------------------------------------------------------------------
make            str18   %-18s                 Make and model
price           int     %8.0gc                Price
mpg             int     %8.0g                 Mileage (mpg)
rep78           int     %8.0g                 Repair record 1978
headroom        float   %6.1f                 Headroom (in.)
trunk           int     %8.0g                 Trunk space (cu. ft.)
weight          int     %8.0gc                Weight (lbs.)
length          int     %8.0g                 Length (in.)
turn            int     %8.0g                 Turn circle (ft.)
displacement    int     %8.0g                 Displacement (cu. in.)
gear_ratio      float   %6.2f                 Gear ratio
foreign         byte    %8.0g      origin     Car origin
-------------------------------------------------------------------------------
Sorted by: foreign
```

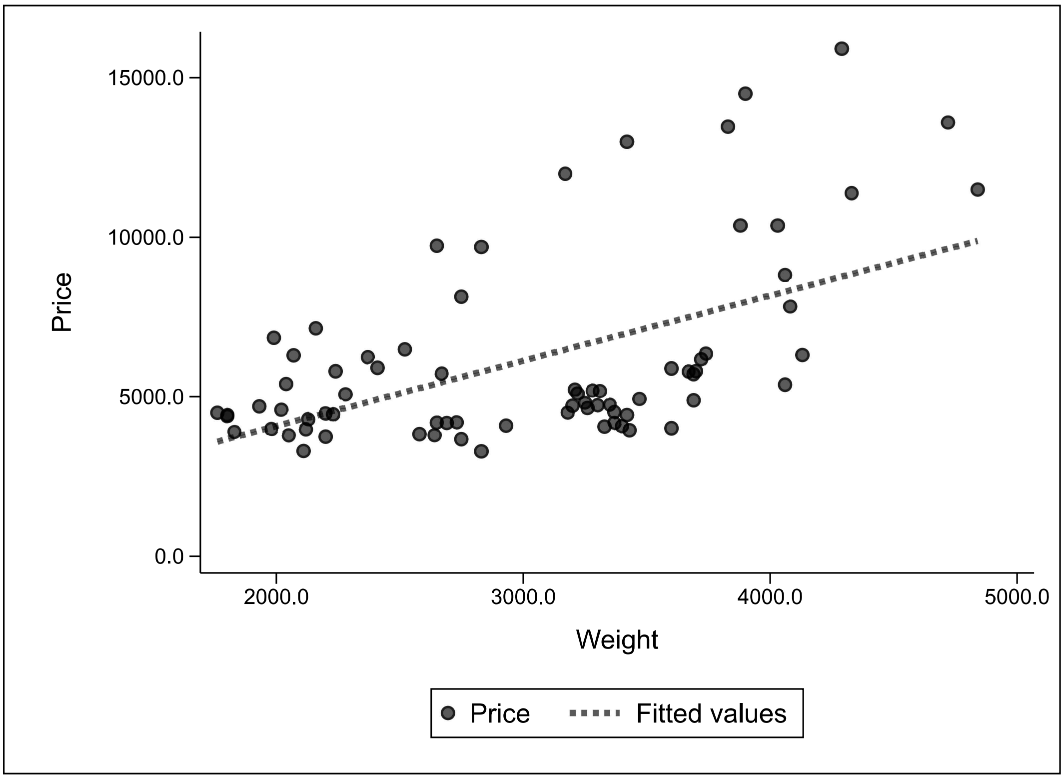
Scatter plot analysis

We can check if an association exists between the price of the car and its weight.

In [7]:
```
%%stata
twoway (scatter price weight, mcolor(black%65)) ///
(lfit price weight, lcolor(black%65) lpattern(shortdash) lwidth(thick)), ///
ytitle(Price, size(medsmall) color(black) ///
margin(medsmall)) ///
xtitle(Weight, size(medsmall) ///
color(black) margin(medsmall)) ///
ylabel(#5, format(%9.1f) angle(horizontal) labsize(small) nogrid) ///
xlabel(, format(%9.1f) labsize(small)) ///
legend(rows(1) symxsize(*0.5)) ///
graphregion(lcolor(black) lwidth(thin) fcolor(white))
```

```
. twoway (scatter price weight, mcolor(black%65)) ///
> (lfit price weight, lcolor(black%65) lpattern(shortdash) lwidth(thick)), ///
> ytitle(Price, size(medsmall) color(black) ///
> margin(medsmall)) ///
> xtitle(Weight, size(medsmall) ///
> color(black) margin(medsmall)) ///
> ylabel(#5, format(%9.1f) angle(horizontal) labsize(small) nogrid) ///
> xlabel(, format(%9.1f) labsize(small)) ///
> legend(rows(1) symxsize(*0.5)) ///
> graphregion(lcolor(black) lwidth(thin) fcolor(white))

.
```



Time to do something fancy.

The code chunk below changes the working directory and loads a dataset on imports and exports from Malawi.

In [8]:
```
%%stata
global path "C:\Users\lucio\OneDrive\Desktop\Jupiter\01_tutorials"
global raw "${path}\raw"
global intermediate "${path}\intermediate"
global outputs "${path}\outputs"
global final "${path}\final"
```

```
. global path "C:\Users\lucio\OneDrive\Desktop\Jupiter\01_tutorials"

. global raw "${path}\raw"

. global intermediate "${path}\intermediate"

. global outputs "${path}\outputs"

. global final "${path}\final"

.
```

In [10]:
```
%%stata
use "$final\mwi_trade.dta", clear
describe
```

```
. use "$final\mwi_trade.dta", clear
(Trade statistics for Malawi. 1964 - 2022)

. describe

Contains data from C:\Users\lucio\OneDrive\Desktop\Jupiter\01_tutorials\final\m
> wi_trade.dta
 Observations:            59                  Trade statistics for Malawi.
                                                1964 - 2022
    Variables:             5                  14 Jun 2025 14:05
--------------------------------------------------------------------------------
Variable      Storage   Display    Value
    name         type    format    label      Variable label
--------------------------------------------------------------------------------
t               float   %ty
m               double  %8.0g                 Imports, Millions Malawi Kwacha
x               long    %8.0g                 Exports, Millions Malawi Kwacha
lm              float   %9.0g                 Log of imports
lx              float   %9.0g                 Log of exports
--------------------------------------------------------------------------------
Sorted by: t

.
```

We can investigate the stationarity properties of these two variables: the log of imports and the log of exports, using the traditional Phillips-Perron Unit root test.

In [16]:
```
%%stata
varsoc lm
```

```
Lag-order selection criteria

   Sample: 1968 thru 2022                                    Number of obs = 55
  +---------------------------------------------------------------------------+
  | Lag |    LL       LR      df    p      FPE       AIC       HQIC      SBIC    |
  |-----+---------------------------------------------------------------------|
  |   0 |  -148.15                        13.2732   5.42362   5.43774   5.46012  |
  |   1 |  9.66102  315.62*    1  0.000 .044315* -.278583* -.250355* -.205589* |
  |   2 |  10.3043  1.2866     1  0.257 .044897  -.265612  -.223271  -.156121  |
  |   3 |  10.3332   .0577     1  0.810 .046518  -.230298  -.173843   -.08431  |
  |   4 |  10.6697  .67298     1  0.412 .047666   -.20617  -.135602  -.023685  |
  +---------------------------------------------------------------------------+
   * optimal lag
   Endogenous: lm
    Exogenous: _cons
```

In [18]:
```
%%stata
pperron lm, lags(1)
```

```
Phillips–Perron test for unit root        Number of obs    = 58
Variable: lm                              Newey–West lags =  1

H0: Random walk without drift, d = 0

                                        Dickey–Fuller
                   Test      -------- critical value ---------
              statistic           1%           5%          10%
--------------------------------------------------------------
 Z(rho)           0.158      -19.044      -13.364      -10.748
 Z(t)             0.402       -3.569       -2.924       -2.597
--------------------------------------------------------------
MacKinnon approximate p-value for Z(t) = 0.9816.
```

In [19]:

```
%%stata
varsoc lx
```

```
Lag-order selection criteria

   Sample: 1968 thru 2022                                      Number of obs = 55
  +---------------------------------------------------------------------------+
  | Lag |    LL        LR       df     p       FPE        AIC       HQIC       SBIC    |
  |-----+---------------------------------------------------------------------|
  |   0 |  -144.52                          11.6319   5.29162   5.30573   5.32812  |
  |   1 |  .295846   289.63     1   0.000 .062295   .061969   .090197   .134963* |
  |   2 |    2.252   3.9123*    1   0.048 .060171*    .0272*  .069541*  .136691  |
  |   3 |  2.76659   1.0292     1   0.310 .061251   .044851   .101306   .190839  |
  |   4 |  3.81224   2.0913     1   0.148 .061165   .043191    .11376   .225676  |
  +---------------------------------------------------------------------------+
   * optimal lag
   Endogenous: lx
    Exogenous: _cons
```

In [20]:

```
%%stata
pperron lx, lags(2)
```

```
Phillips–Perron test for unit root        Number of obs    = 58
Variable: lx                              Newey–West lags =  2

H0: Random walk without drift, d = 0

                                        Dickey–Fuller
                   Test      -------- critical value ---------
              statistic           1%           5%          10%
--------------------------------------------------------------
 Z(rho)           0.010      -19.044      -13.364      -10.748
 Z(t)             0.026       -3.569       -2.924       -2.597
--------------------------------------------------------------
MacKinnon approximate p-value for Z(t) = 0.9606.
```

We stop here for now!

Have you tried this setup? I would be happy to learn more. I will be exploring this integration further.

In [ ]: