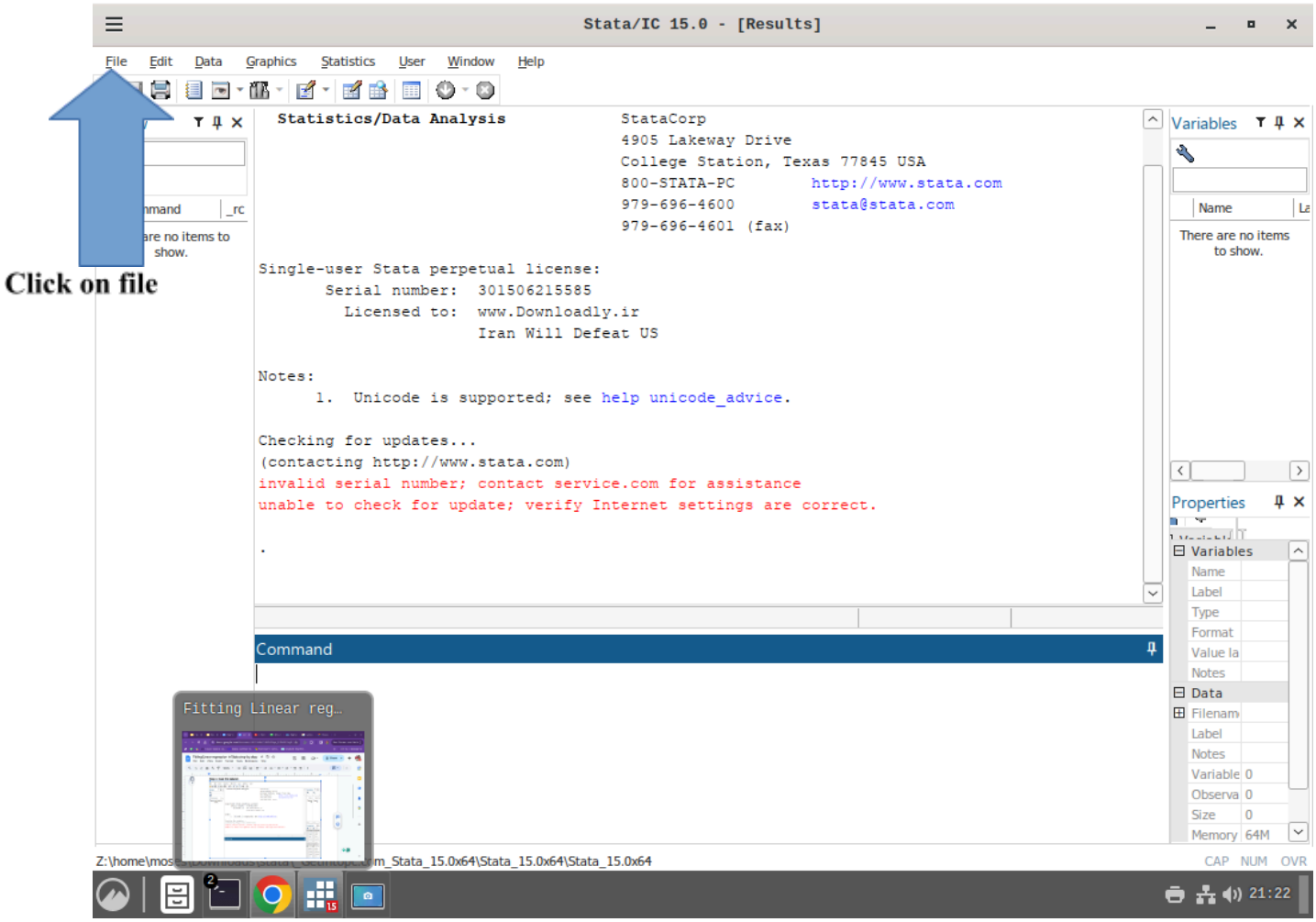
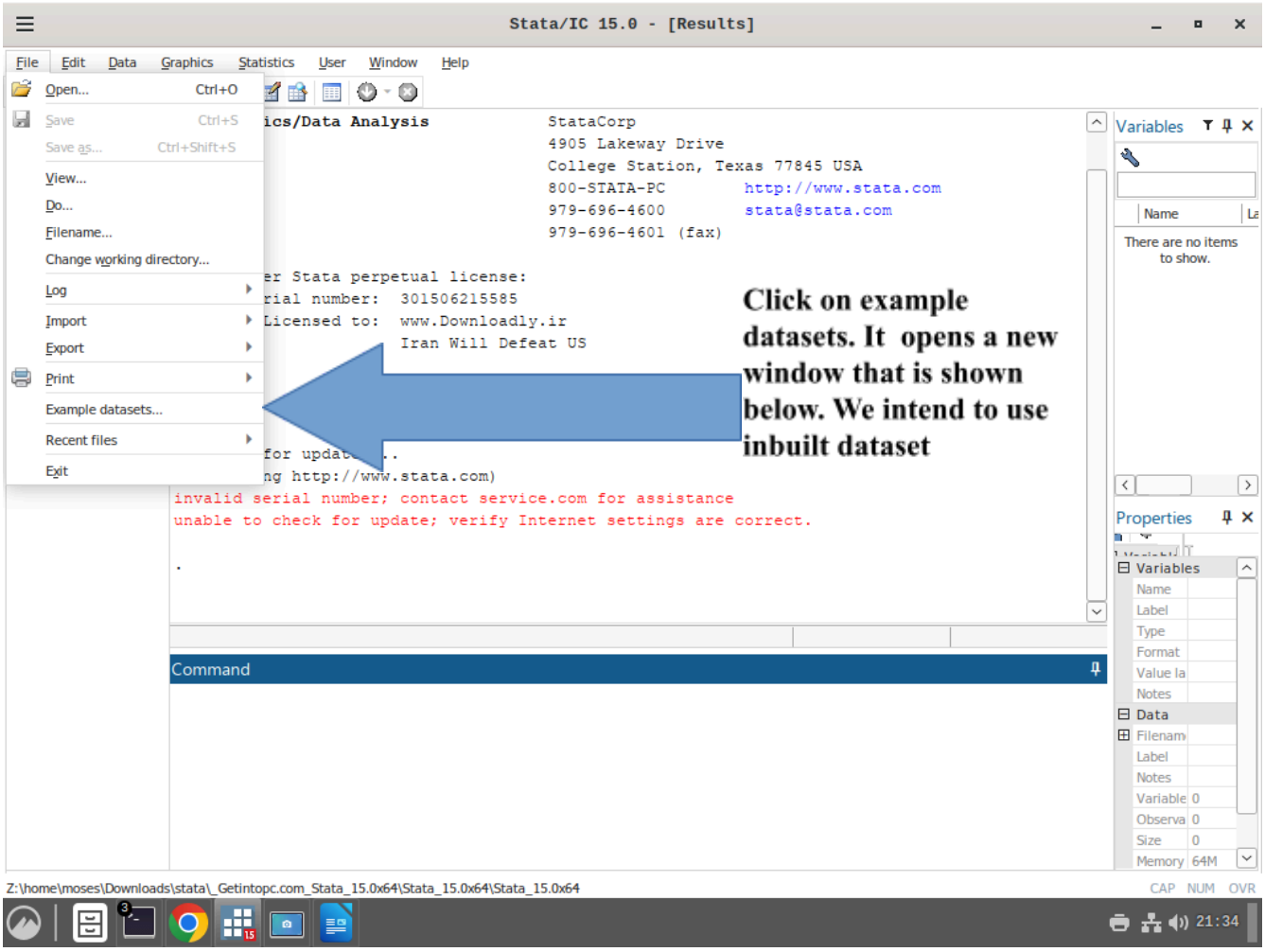
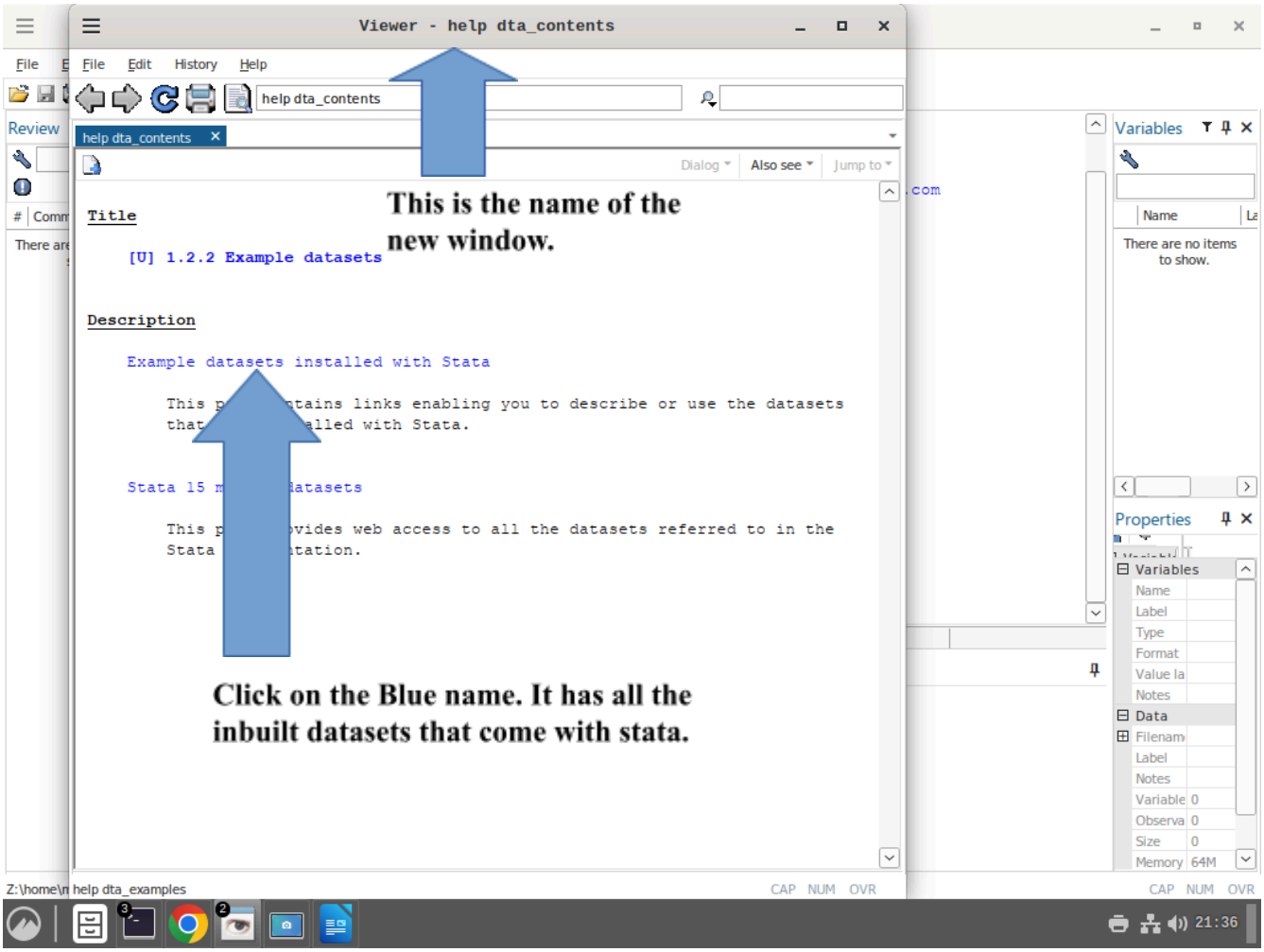
Step 1: load the data



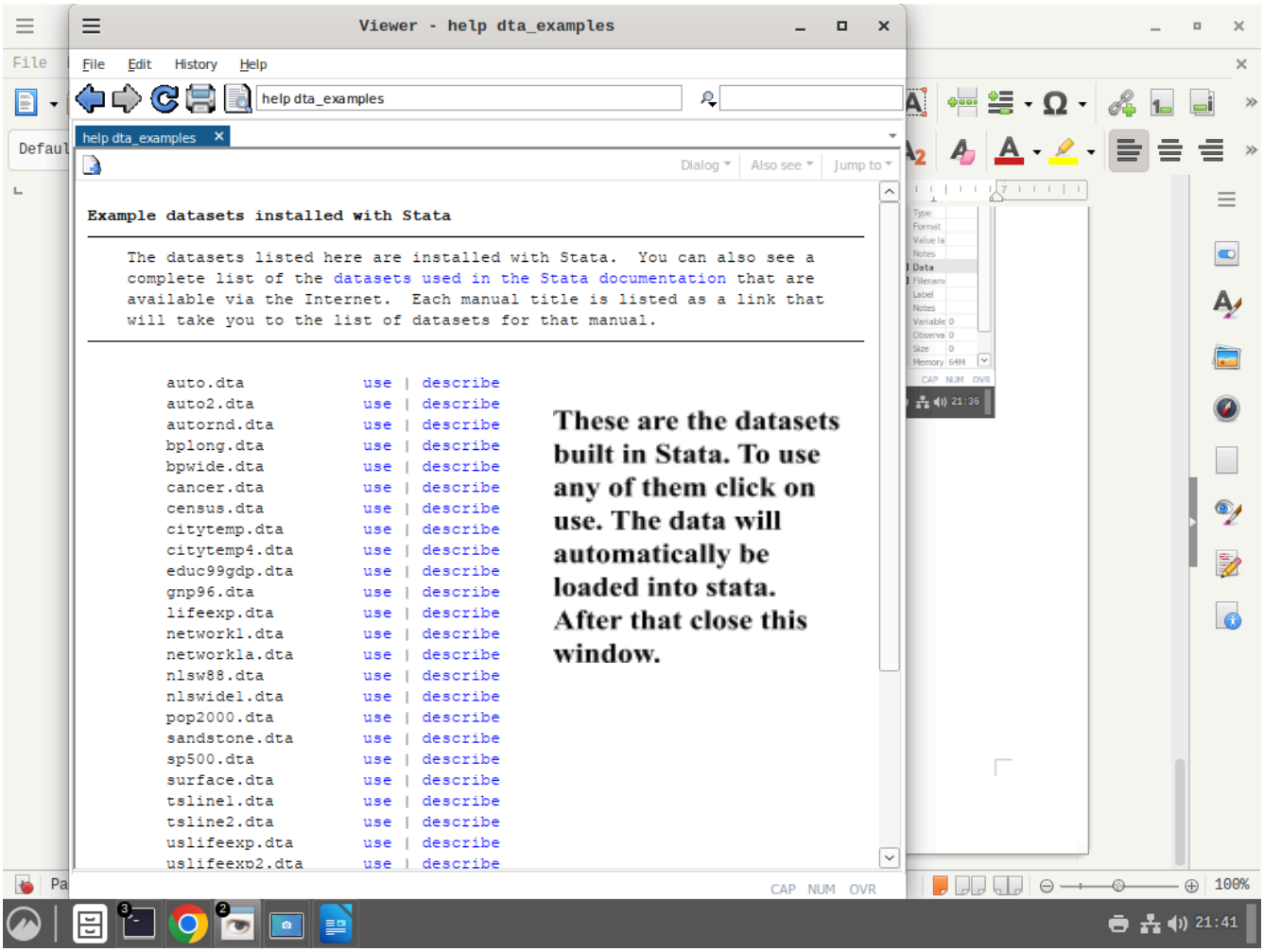
Now we have to get the inbuilt dataset



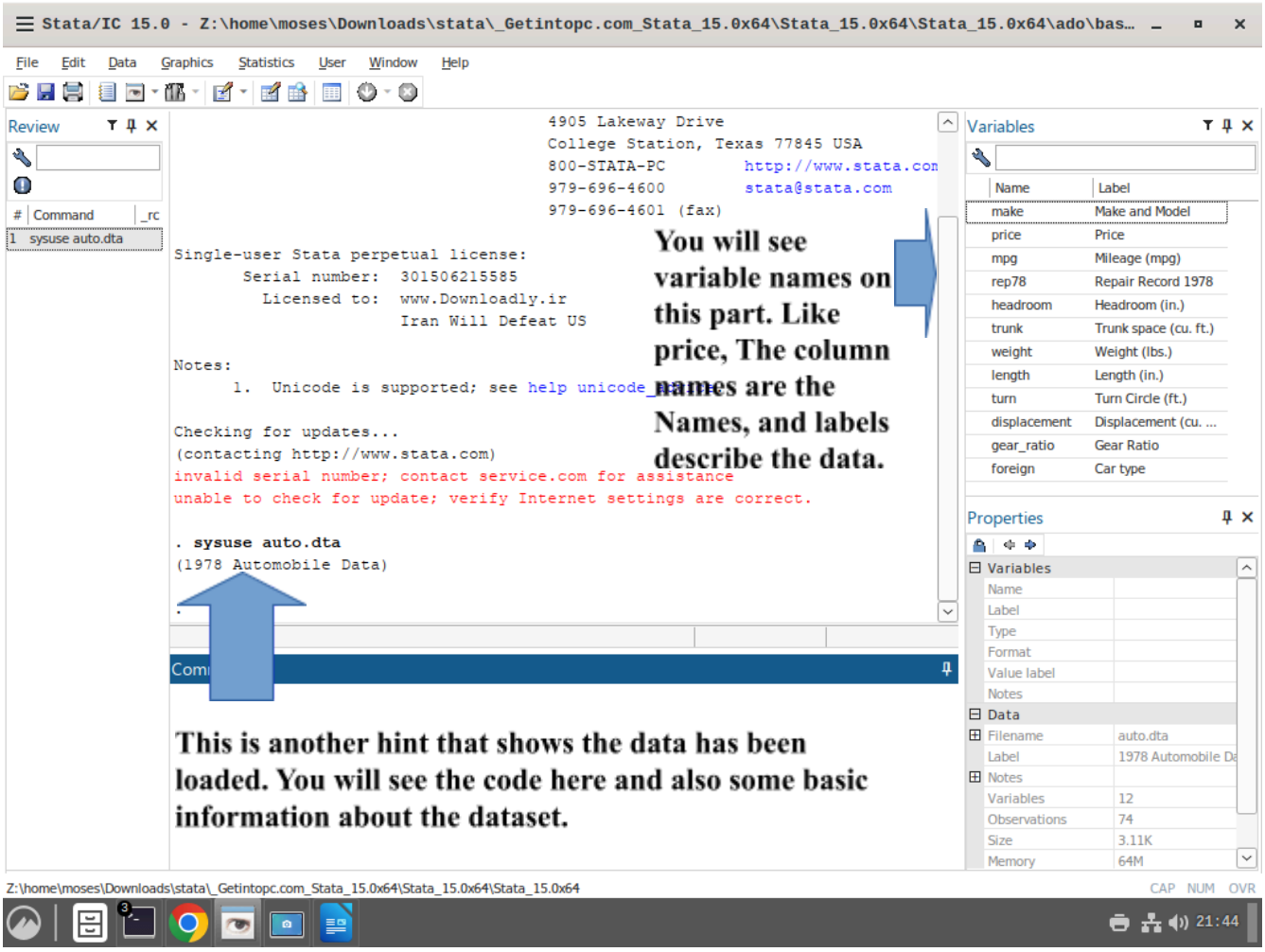
Lets go to the new windows that opens up after this



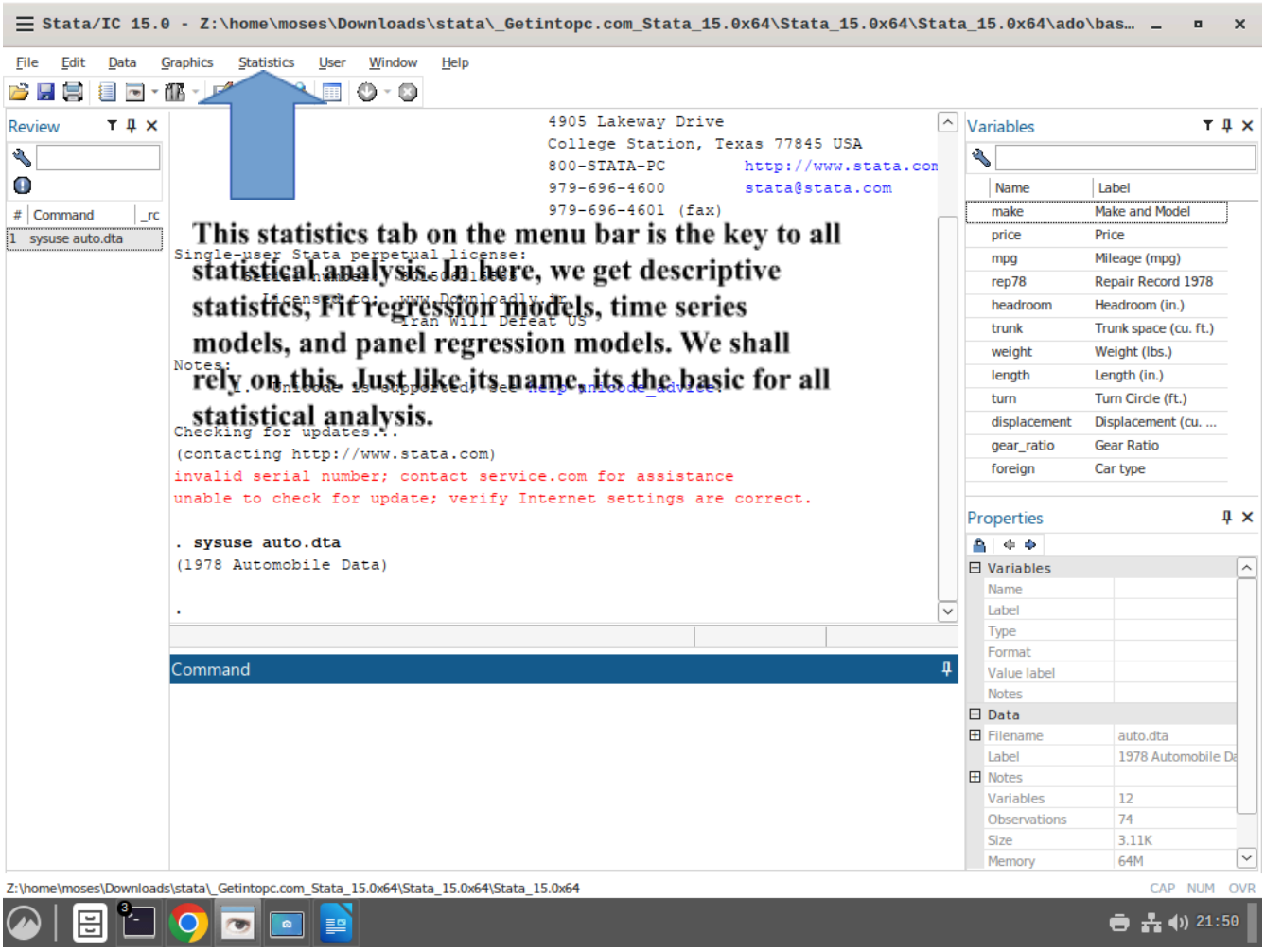
Let's view the datasets



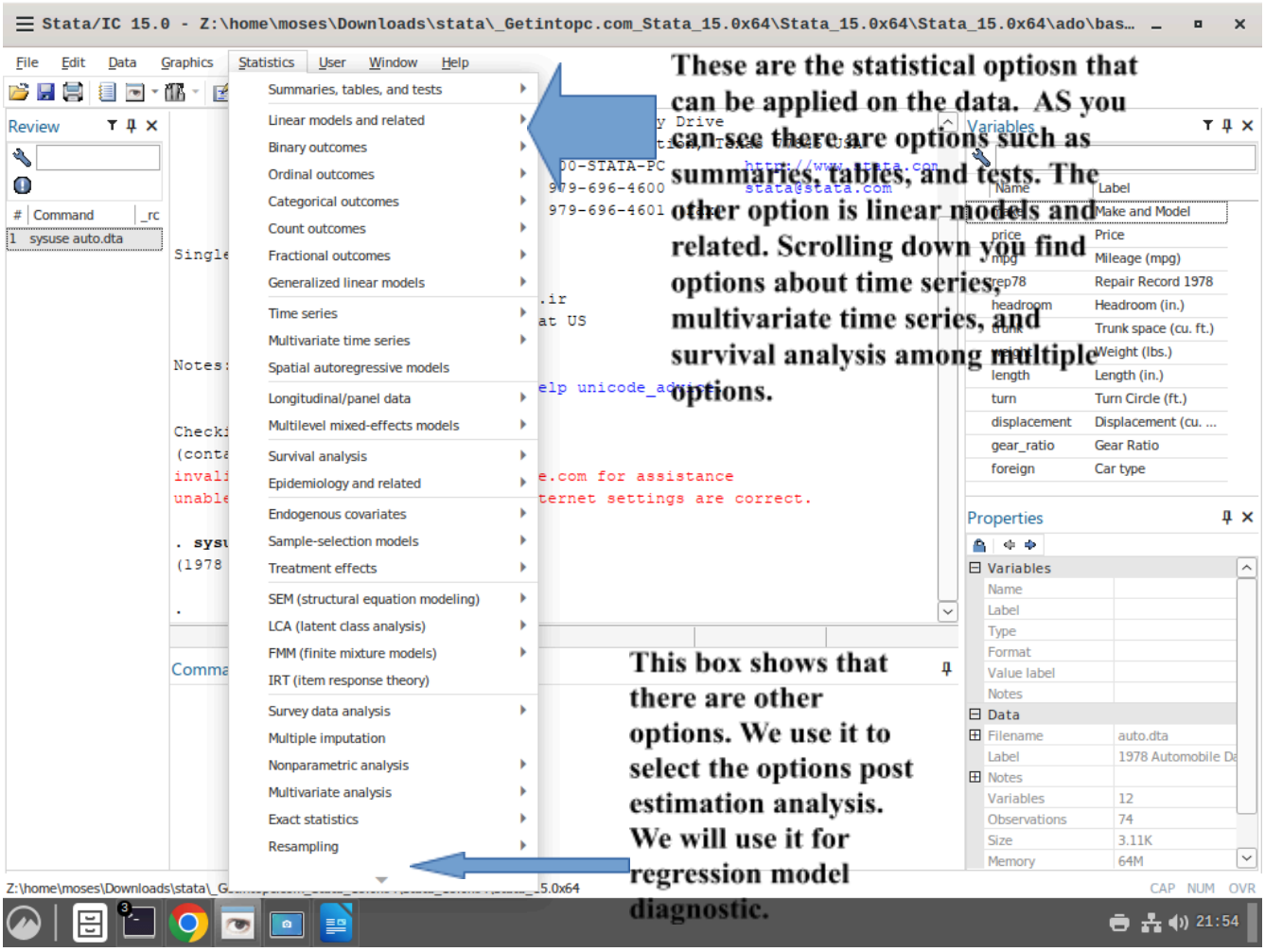
How do we know that the dataset has been loaded? We selected auto.dta



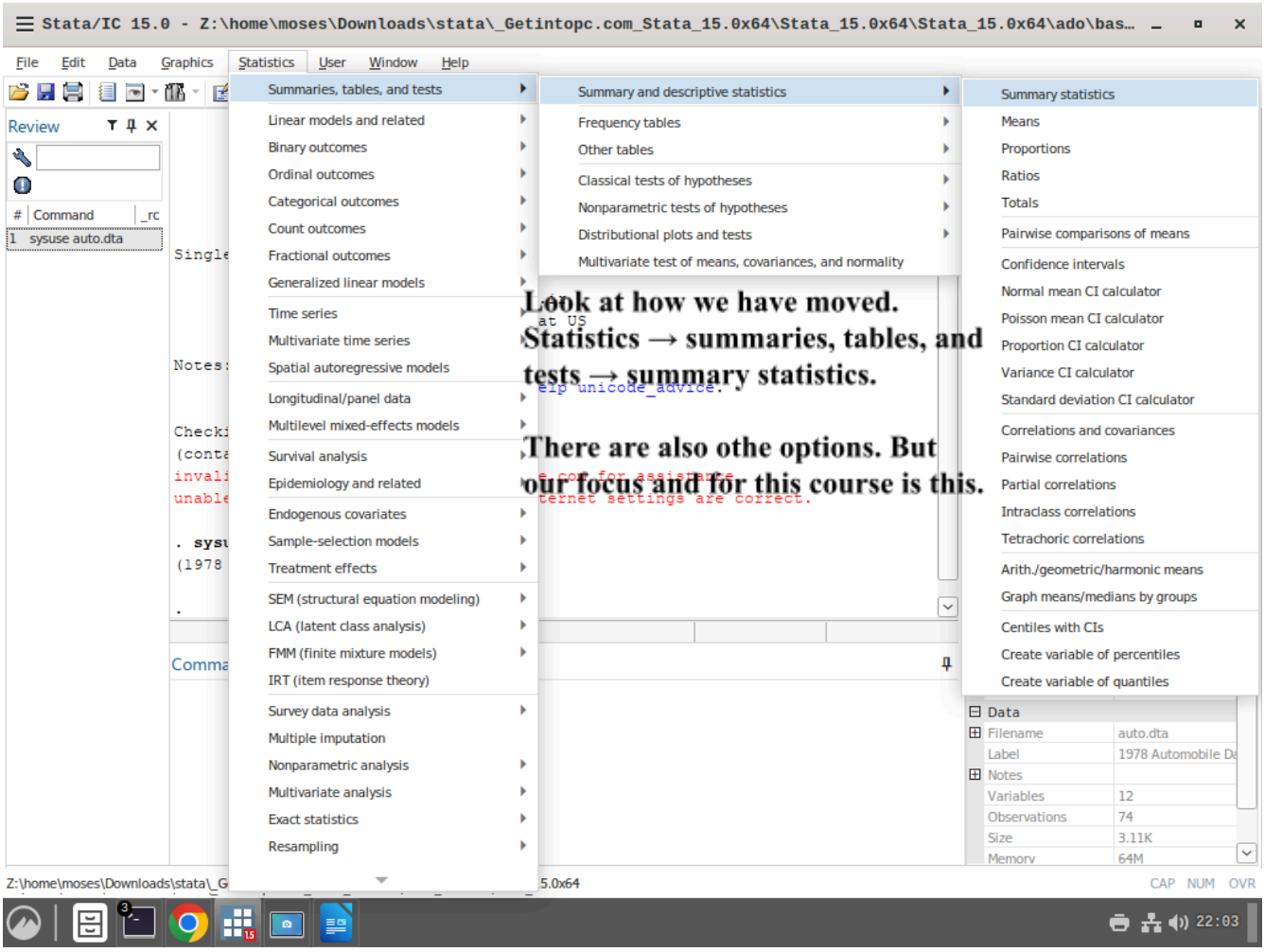
We have the dataset. Lets get the basic descriptive statistics



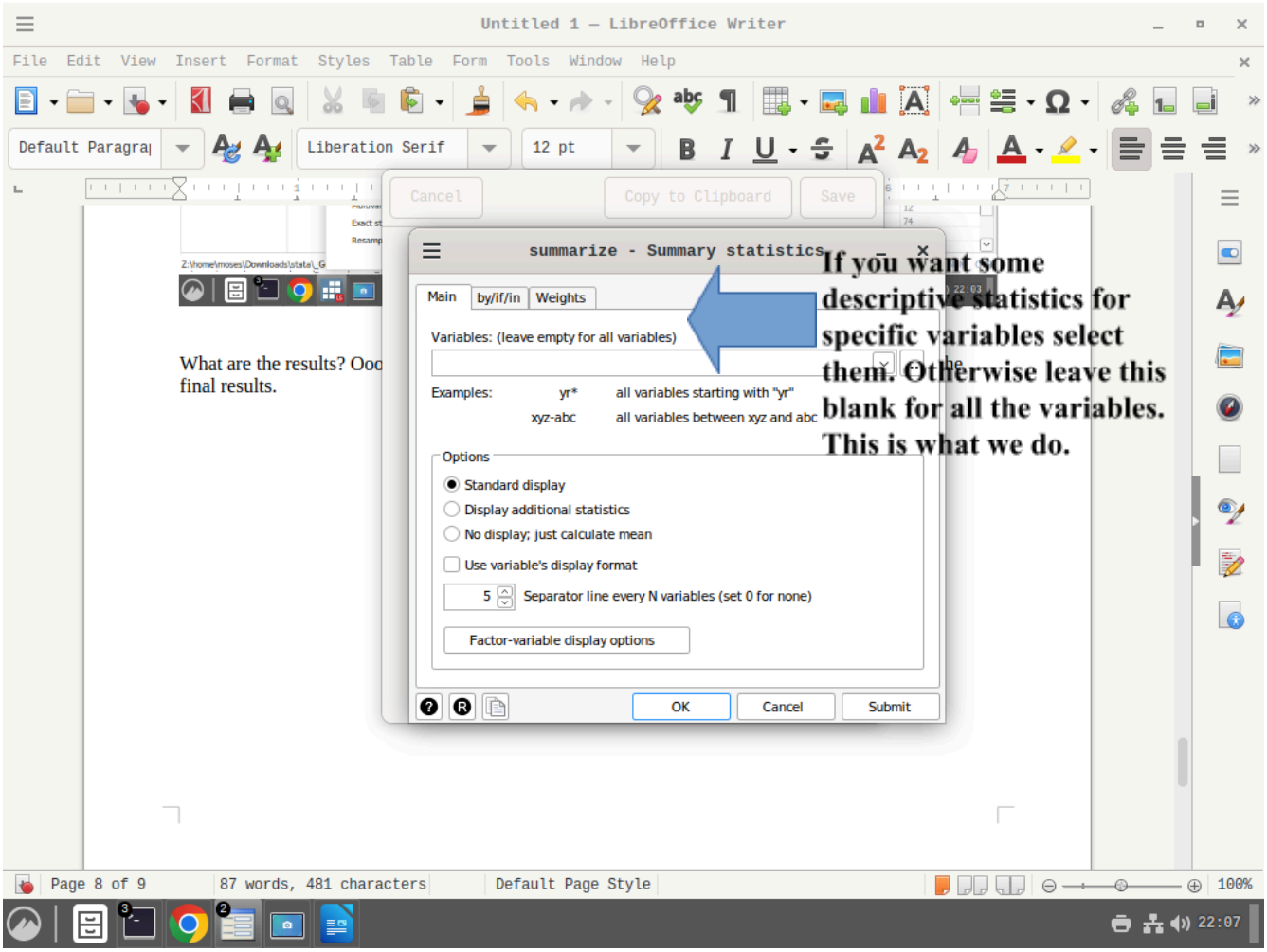
Lets go deeper



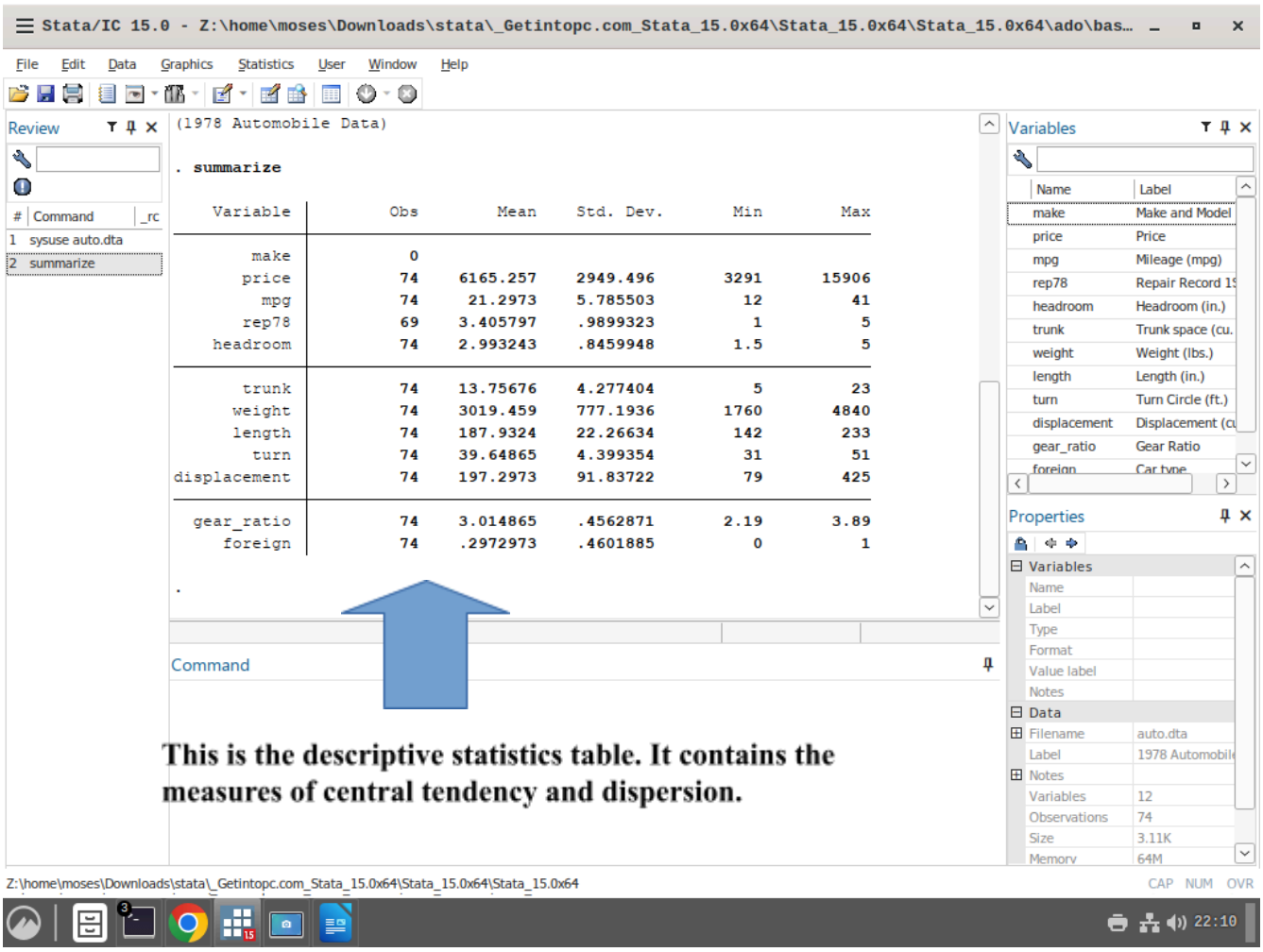
lets do descriptive statistics. We select the first options summaries, tables and tests.



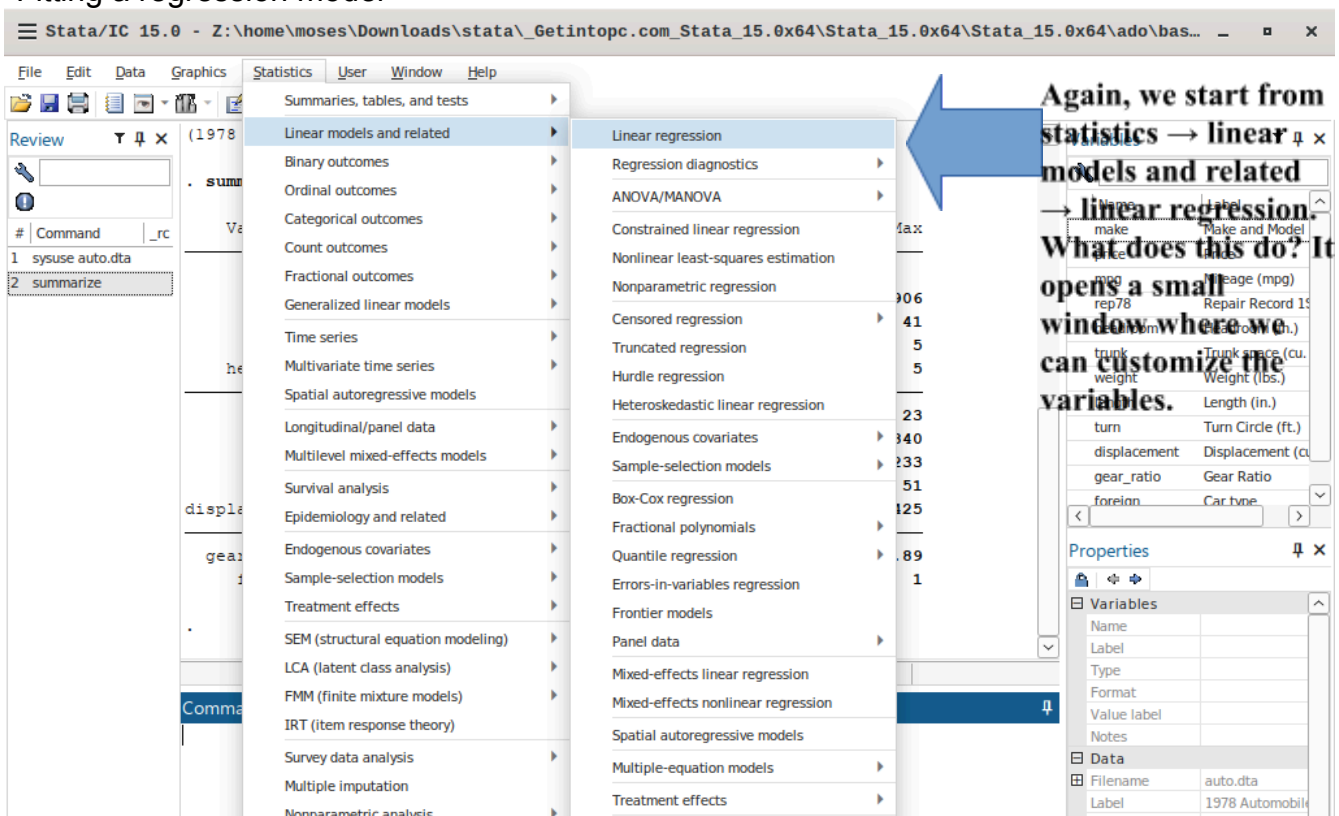
What are the results? Ooops! Not the output but a new small window that you customize before the final results.



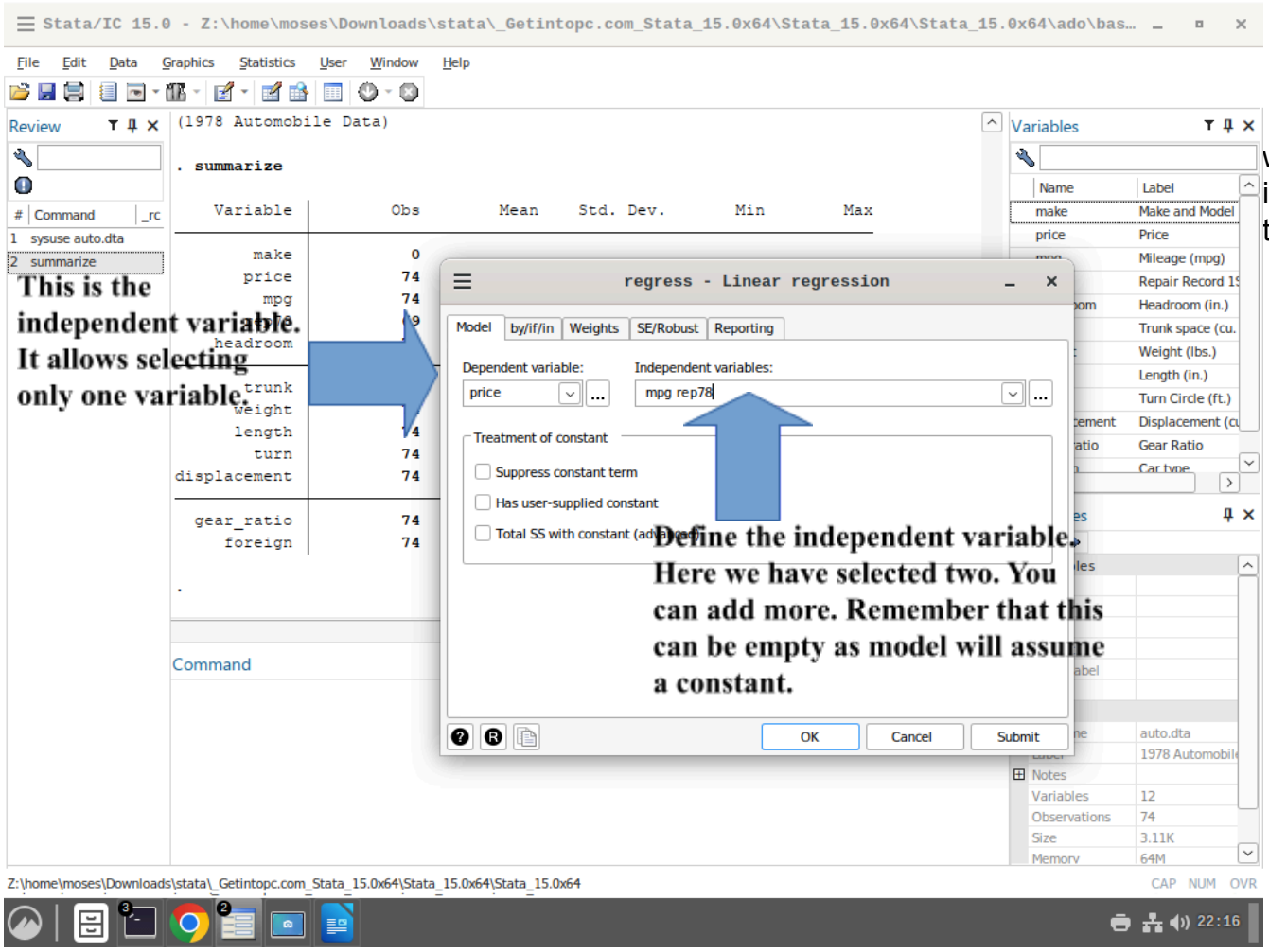
The descriptive statistics table.



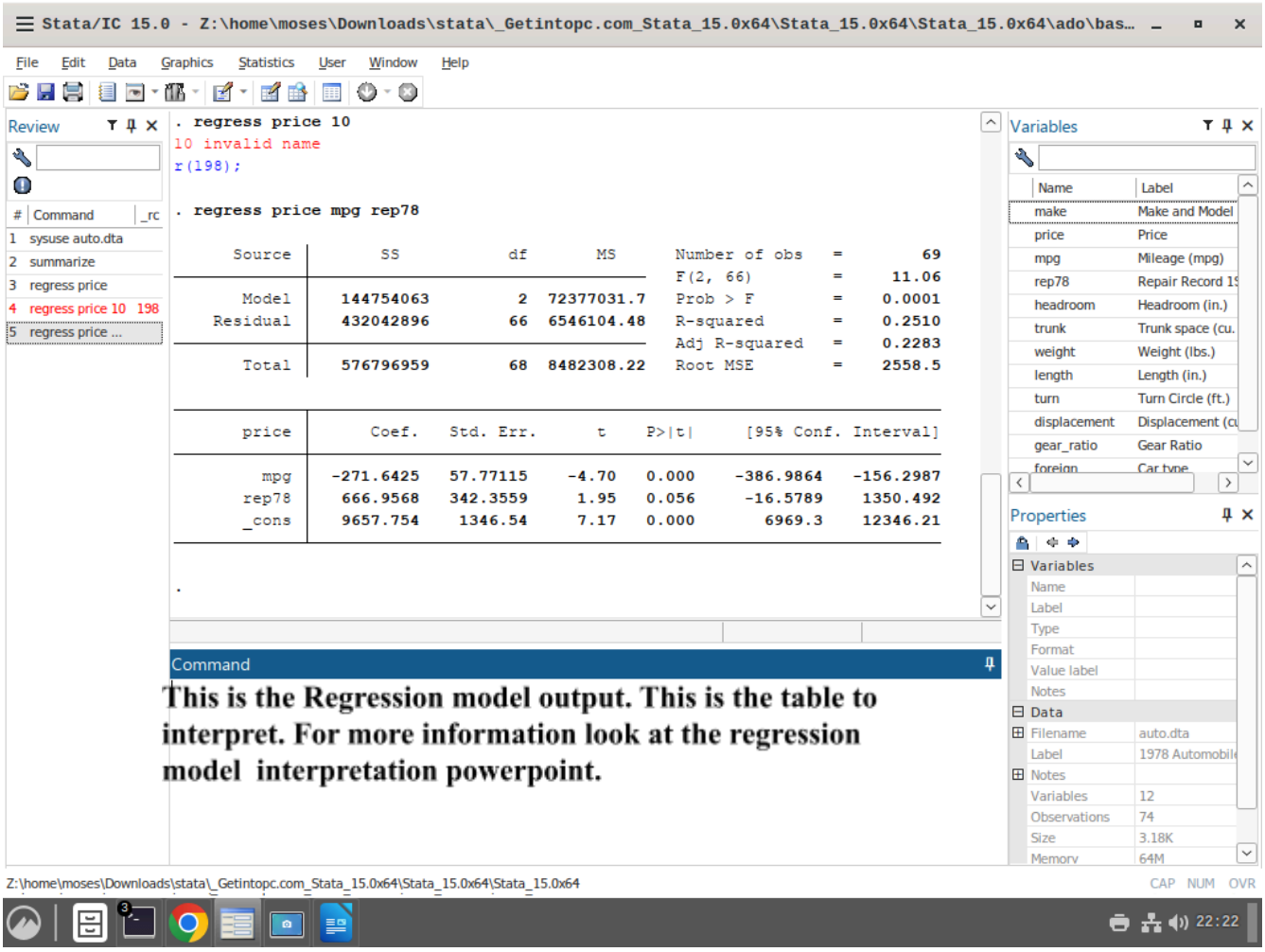
Fitting a regression model



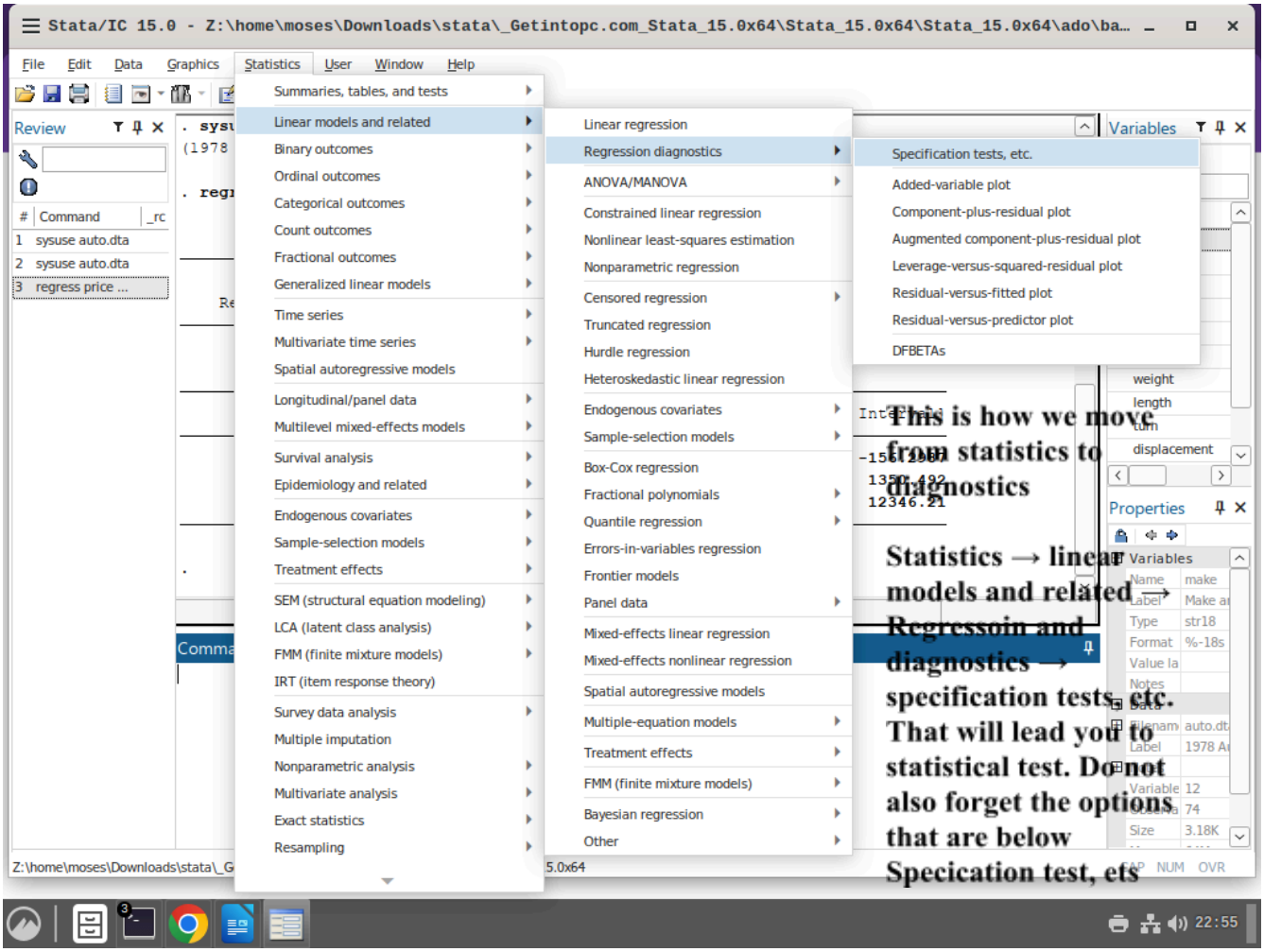
Defining the regression model



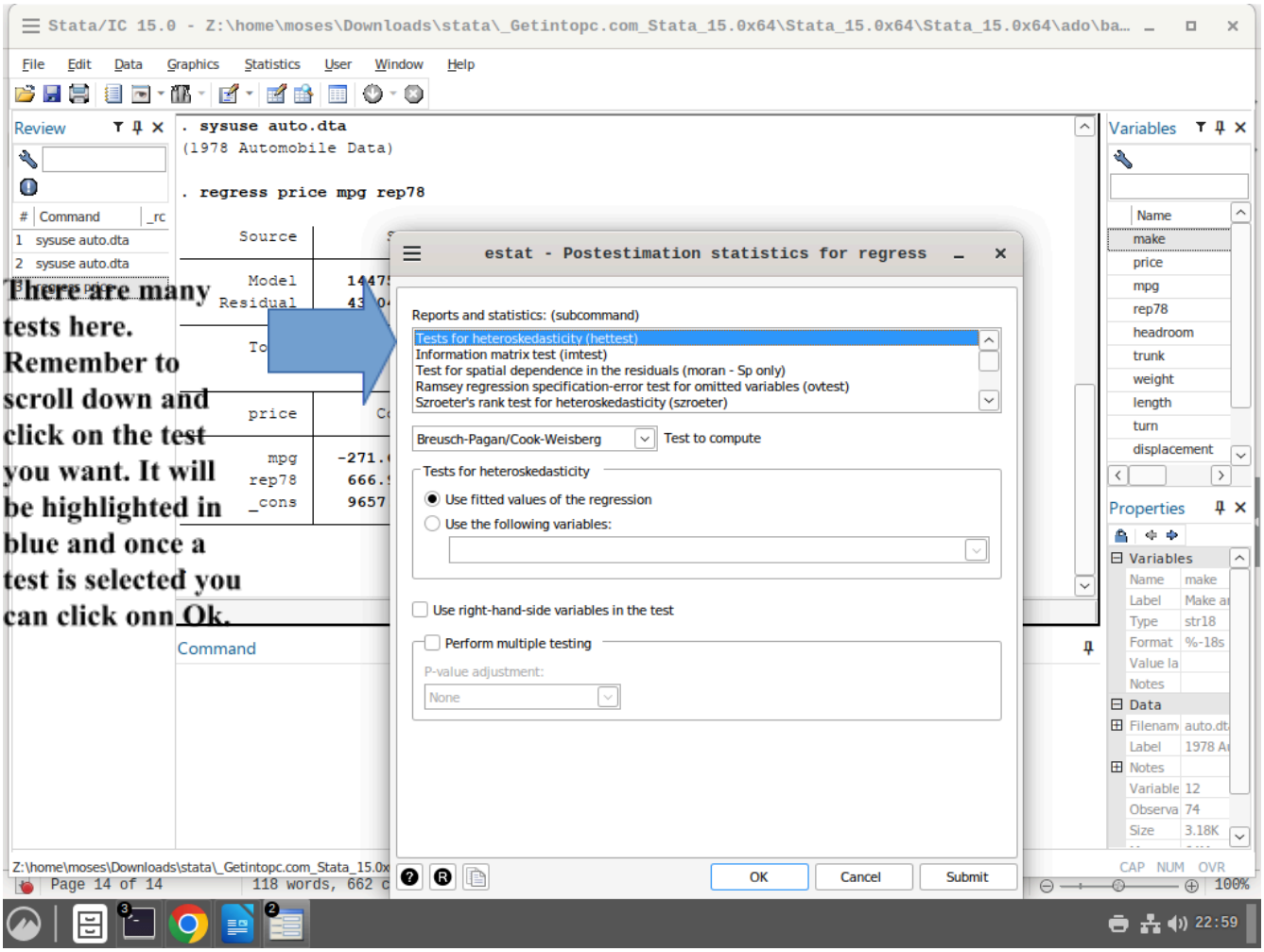
what is the ouput?



Lets go to model diagnostics



lets click on the last part and see the results



Lets end with a summary of the diagnostics tests

In this chapter, we have used a number of tools in Stata for determining whether our data meets the regression assumptions. Below, we list the major commands we demonstrated organized according to the assumption the command was shown to test.

- **Detecting Unusual and Influential Data**
    - **predict** — used to create predicted values, residuals, and measures of influence.
    - **rvpplot** — graphs a residual-versus-predictor plot.

    - **rvfplot —** graphs residual-versus-fitted plot.
    - **lvr2plot** — graphs a leverage-versus-squared-residual plot.
    - **dfbeta —** calculates DFBETAs for all the independent variables in the linear model.
    - **avplot** — graphs an added-variable plot, a.k.a. partial regression plot.
- **Tests for Normality of Residuals**
    - **kdensity** — produces kernel density plot with normal distribution overlayed.
    - **pnorm** — graphs a standardized normal probability (P-P) plot.
    - **qnorm** — plots the quantiles of varname against the quantiles of a normal distribution.
    - **iqr** — resistant normality check and outlier identification.
    - **swilk** — performs the Shapiro-Wilk W test for normality.
- **Tests for Heteroscedasticity**
    - **rvfplot —** graphs residual-versus-fitted plot.
    - **hettest** — performs Cook and Weisberg test for heteroscedasticity.
    - **whitetst** — computes the White general test for Heteroscedasticity.
- **Tests for Multicollinearity**
    - **vif** — calculates the variance inflation factor for the independent variables in the linear model.

        - **collin** — calculates the variance inflation factor and other multicollinearity diagnostics
- **Tests for Non-Linearity**
        - **acprplot** — graphs an augmented component-plus-residual plot.
        - **cprplot** — graphs component-plus-residual plot, a.k.a. residual plot.
- **Tests for Model Specification**
        - **linktest** — performs a link test for model specification.
        - **ovtest** — performs regression specification error test (RESET) for omitted variables.

Questions

1. In the same manner that this analysis has used inbuilt dataset, you are going to use the inbuilt dataset nlsw88.dta . Create a linear regression model with your dependent variable and independent variable. Interpret the model and do model diagnostics.
2. Using the laptop.csv dataset, use variables inches, RAM (GB), and Weight (kg) as the independent variables and Price as the dependent variables to model a linear regression model. Ensure that you plot graphs, and do diagnostic analysis.