Georgetown CSSA Data Workshop #2

Date: 2023/3/2
Shiyao Wang

Software:

1. *Stata (GU WebStore and any versions are good to go)
   Please refer to the notice in the WeChat group for more free download options
2. *Microsoft Excel (WPS may as a substitute, but Excel is preferred)

means required for this session

Topic:

1. Replicate some simple Excel operations in Stata
   a. import data: https://grade-data.oss-cn-hangzhou.aliyuncs.com/galaviz_rawdata/auto.csv
   b. browse data
   c. summarize the mean for a column
   d. draw a histogram

2. Replicate some Excel intermediate operations in Stata
   a. date and time
   b. COUNTIF
   c. COUNTIFS
   d. VLOOKUP
   e. SUMIFS
   f. pie chart visualization
   g. sort and replace string values
   h. pivot table
   i. string concatenation
   j. conditional formatting

   Some questions to think about:

   (1) Which one is easier and more efficient, Excel or Stata?
   (2) Can you classification the operations we've conducted into several groups, which can potentially represent the life cycle of data?
   (3) To write Stata codes, what are the potential difficulties (knowing which cmd or syntax itself?)

3. 4 Essential tips/to know for Stata Beginners:
   a. Software Interface vs. Programming
   b. Syntax
   c. Stata talks to Variables
   d. Interfaces: Many separate windows = easier to start with one, but can be hard to master all