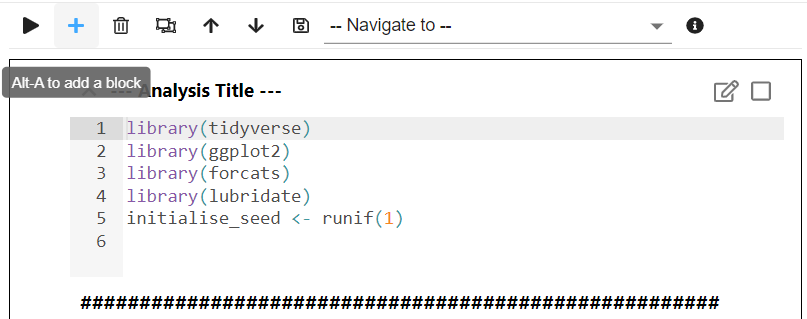
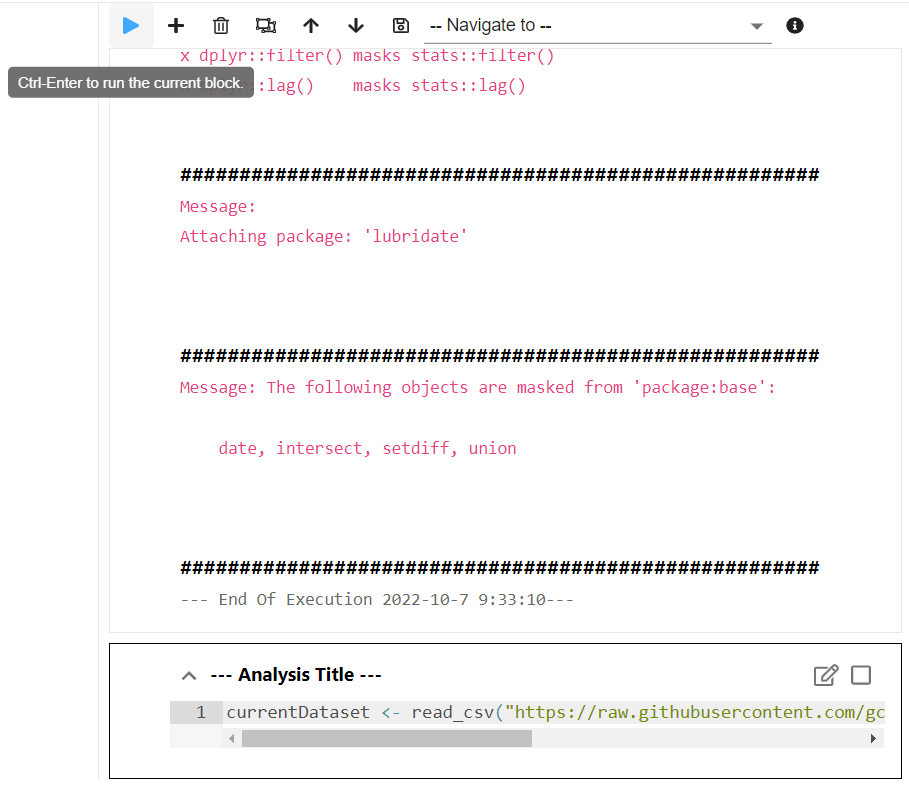
**Matching**

1. Add a new block by click the “+” button on the right panel.

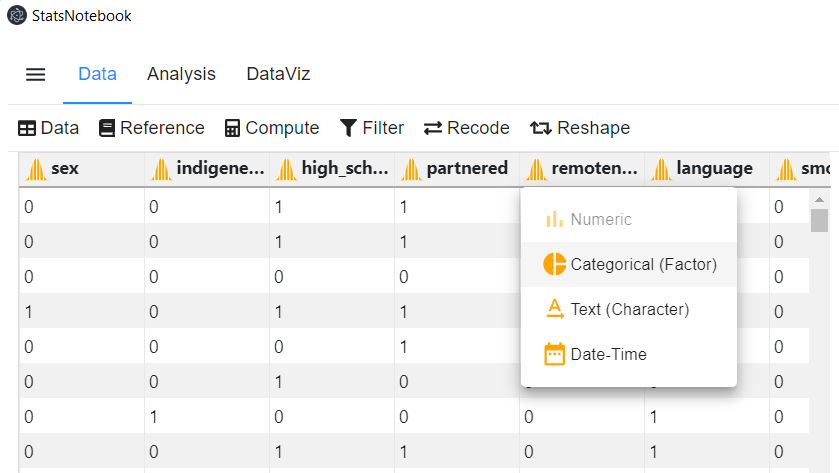


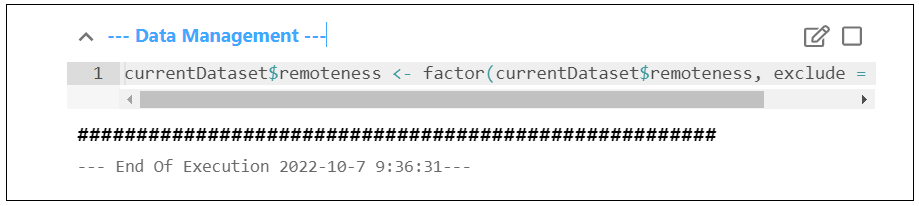
1. To load the data, type the following (or Copy and Paste) into the new block (Please note that the codes below is slightly different from those provided in the appendix to my paper because StatsNotebook always load the data into the variable “currentDataset”)

currentDataset <- read\_csv("https://raw.githubusercontent.com/gckc123/Causal\_Analysis\_Addiction\_Examples/main/smoking\_psyc\_distress.csv")  
  
And then click “Run”.

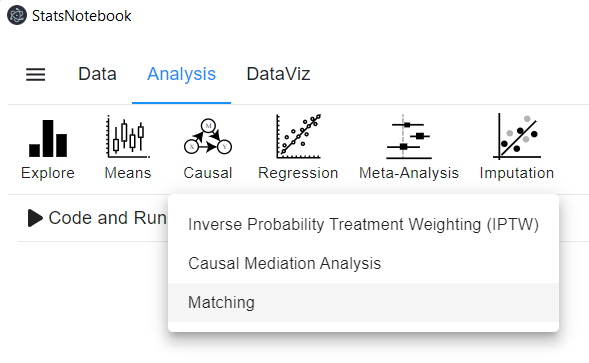


1. Since the variable remoteness has three categories, convert it into a factor variable.
   1. Click on “Data” on the left panel to show the dataset.
   2. Click on the orange icon next to “remoteness” label. A new block with codes will appear on the right panel.



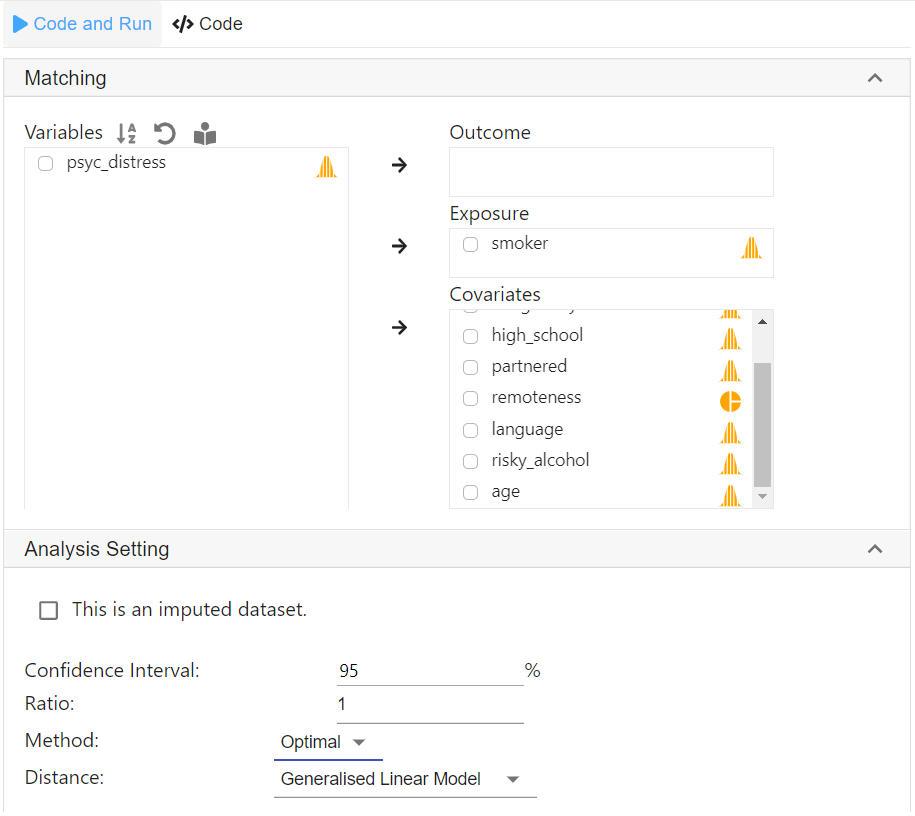


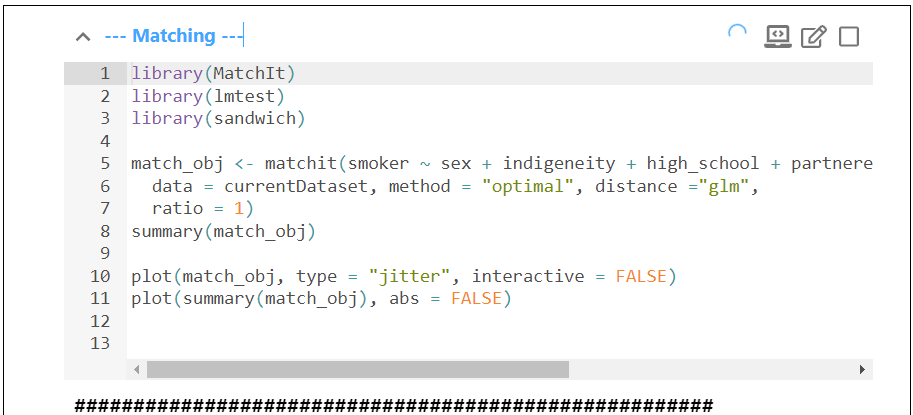
1. Click on “Analysis” and then “Matching”.

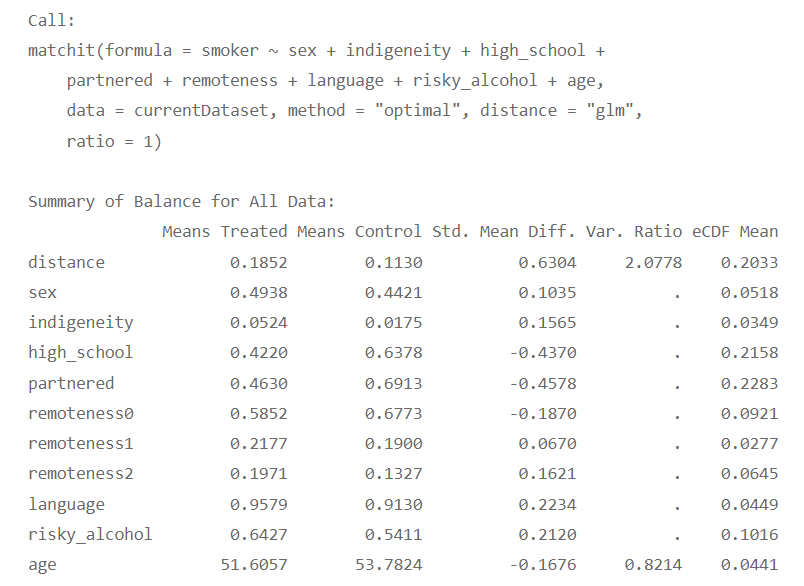
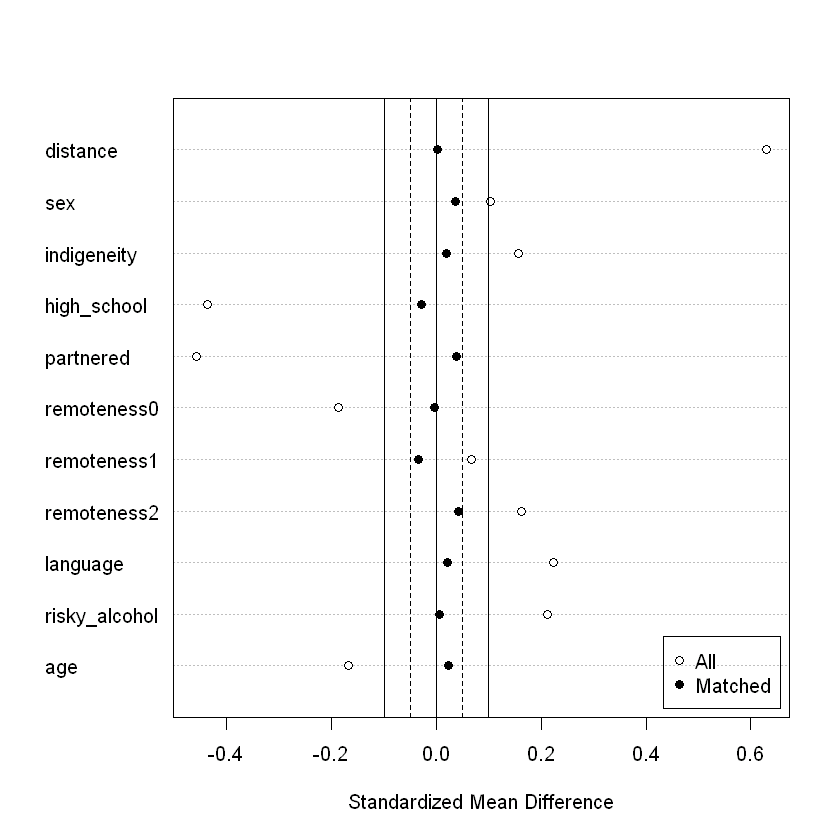


1. To conduct matching,
   1. Select “smoker” as the exposure variable, and “sex”, “indigeneity”, “high\_school”, “partnered”, “remoteness”, “language”, “risky\_alcohol” and “age” as covariates.
   2. Under “Analysis setting”, select “optimal” as the matching method.
   3. Click “Code and Run” at the top.

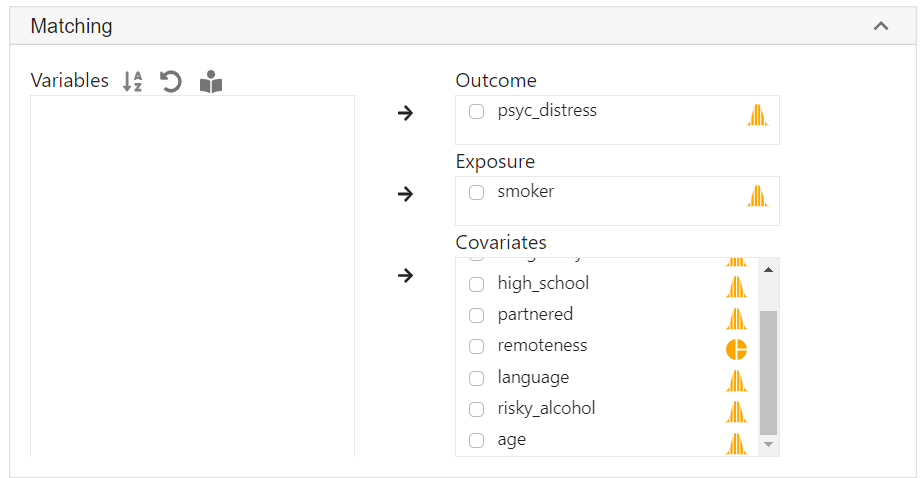
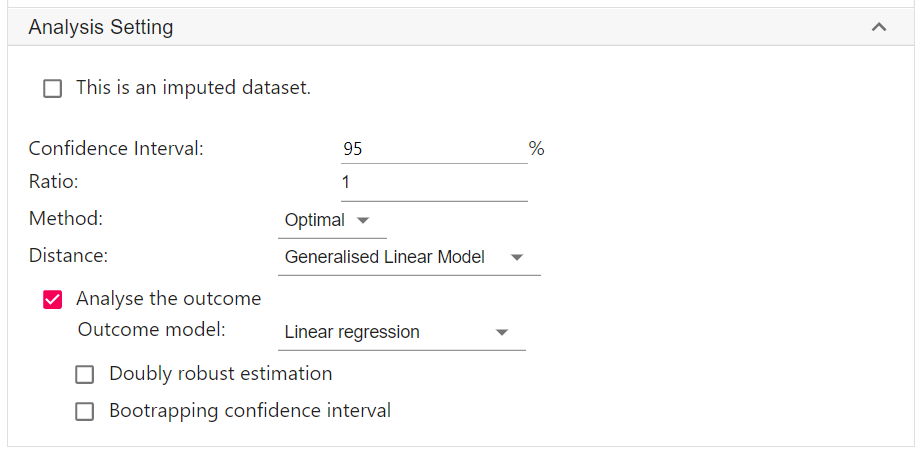
A new block with codes will appear on the right panel, with results from this procedure displayed below the codes.

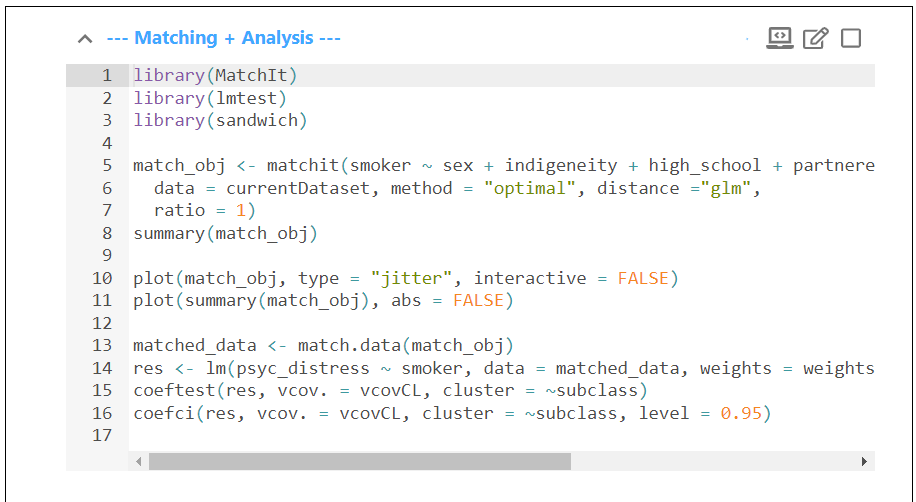
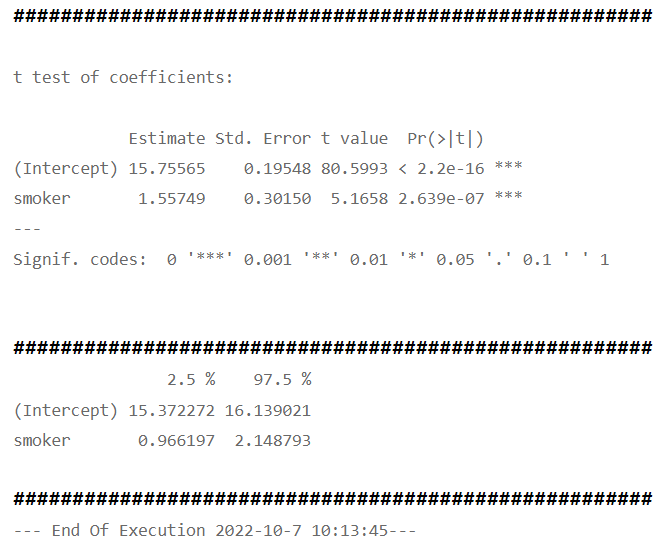




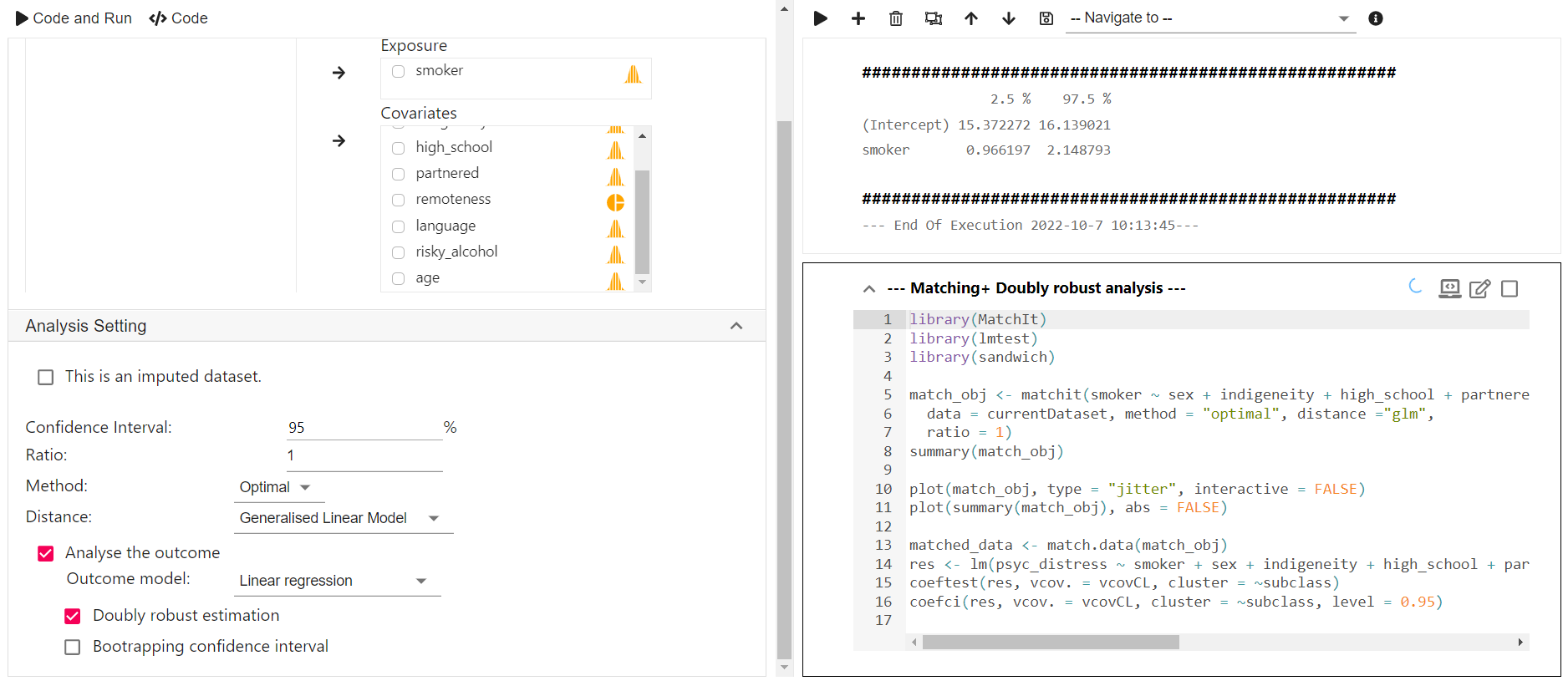
  
 

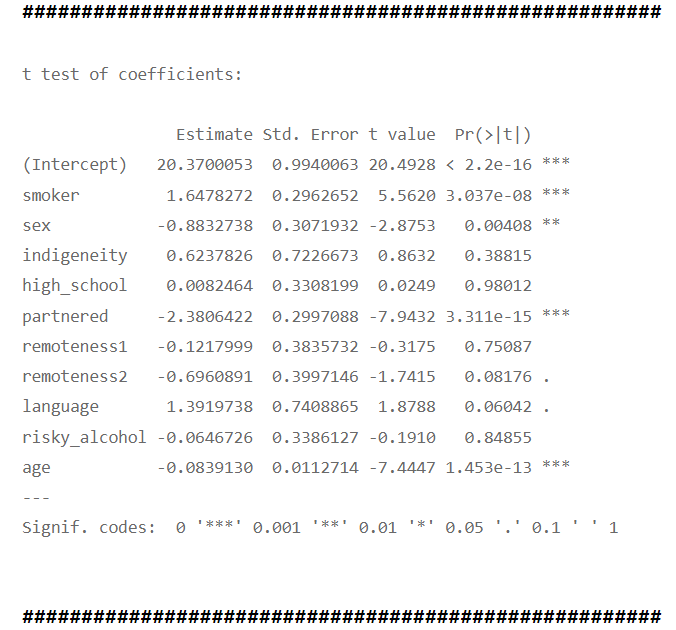
1. Once a satisfory match is obtained, we can conduct the second step – comparing the difference in psychological distress between smoker and non-smoker.
   1. Select “psyc\_distress” as the outcome variable.
   2. In “Analysis setting”, click “Analyse the outcome”. The model will be a linear regression as psyc\_distress is treated as a continuous variable.
   3. Click “Code and Run”.

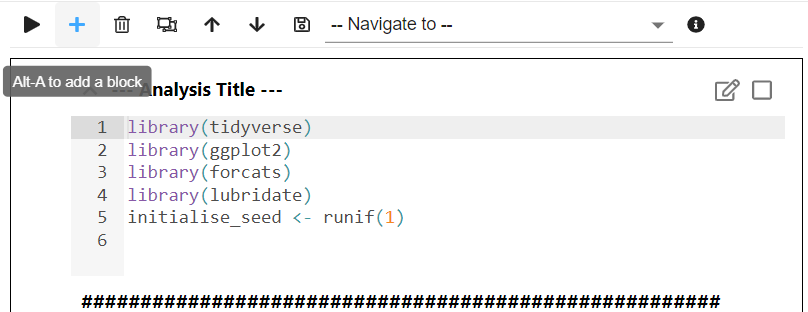
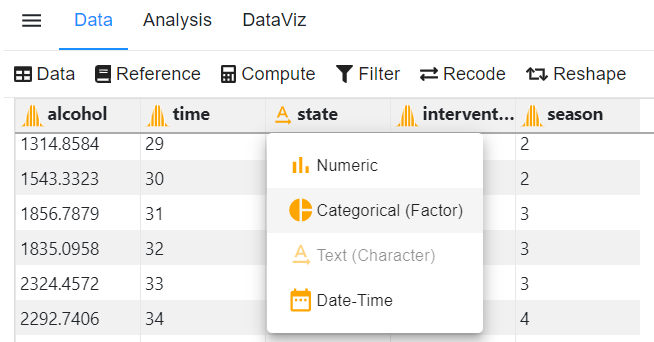
 

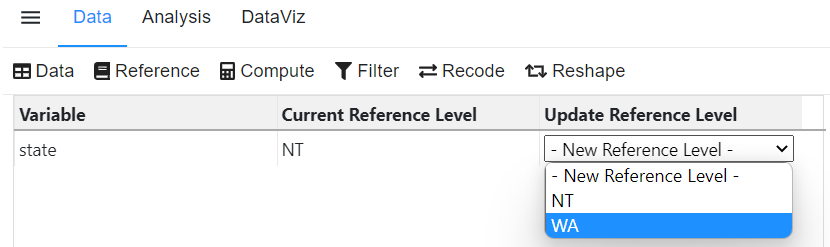
1. To run a doubly robust analysis (This step is not necessary),
   1. click “Doubly robust estimation” under “Analysis setting”.
   2. Click “Code and run”.



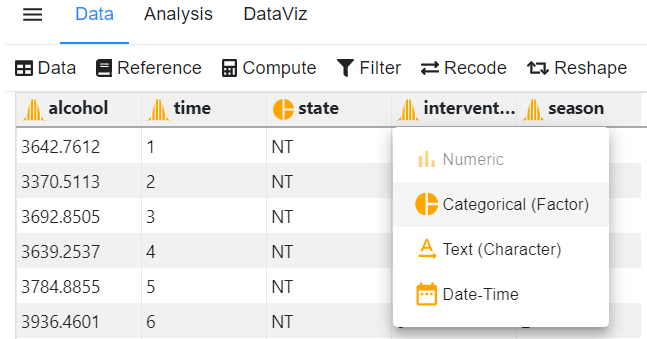


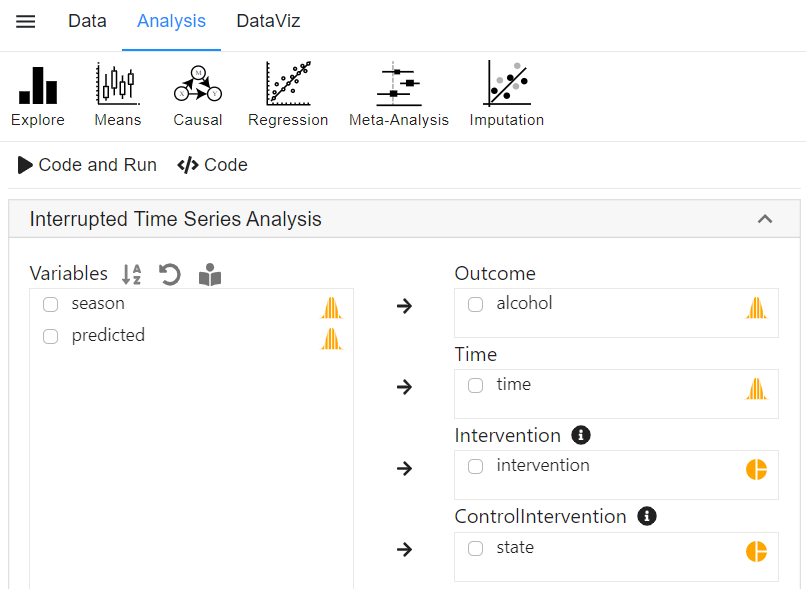
**Interrupted time series analysis**

1. To load the dataset,
   1. Click the “+” icon on the right panel, a new block will appear. 
   2. Type the following code into this new block.  
        
      currentDataset <- read\_csv("https://statsnotebook.io/blog/data\_management/example\_data/alcohol\_data\_NTWA.csv")  
        
      
2. Convert “State” and “intervention” into a factor variable
   1. Click “Data” on the left
   2. Click the orange icon next to “State”
   3. Click “Categorical (Factor)”. A new block with code will appear on the right panel.  
      
   4. For this analysis, we will also use “Western Australia” as the reference.
      1. Click “Reference”, under “Update reference level”, select “WA”.



* 1. Click the orange icon next to “intervention”, and click “Categorical (Factor)”.



1. To conduct an interrupted time series analysis,
   1. Click “Analysis”, “Regression” and then “Interrupted time series analysis”.
   2. Select “alcohol” as Outcome, “time” as Time, “intervention” as Intervention, and “state” as Control Intervention.
   3. In “Analysis Setting”, click “ACF/PACF” for the autocorrelation plot.
   4. To adjust for seasonality, click “Harmonic terms for seasonality adjustment”. We will use the default setting in which we assume that a cycle contains 12 time points (Since we have monthly data, this means that we assume that there is a yearly cycle).
   5. Click “Code and Run”.  
      
2. The above procedure will generate R codes for the interrupted time series analysis.  
   