

# User Guide for DHIIVAT: Disease History & Income Inequality Visual Analytics Tool

## 1 Menu

Open the DHIIVAT toolkit, there are four main menus as below pictures.



Figure 1 Main page

In the first page of top bar, we have a simple introduction about our project which mentions the motivation of doing this topic. And after clicking on the second page, you can start the steps of data exploring.

## 2 EDA Tab

On the sidebar, we have prepared four topics about individual income and health and five topics about individual control variables. You can choose any interesting topic you want to observe.



Figure 2 EDA tab list

We can click error bar topic first, then, one control panel and one main panel show up. You can click only one province to see its confidence interval in five years (square 1), and there are two standards of confidence interval, 95% and 99%, can be chosen at the drop-down button (square 2).

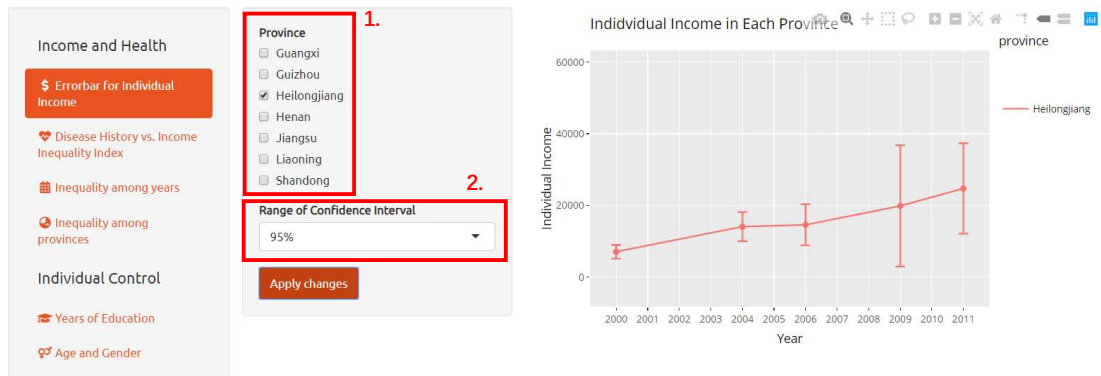


Figure 3 Individual Income comparing panel-1

Another way to use this control panel is clicking two provinces at the same time and this can compare average income in two provinces (square 3). However, we don't suggest comparing more than two provinces at the same time, because the chart will be a mass and confuse you a lot. And after setting up all details, please click on **Apply Changes** button (square 4) and the app will run all settings and give you the result rapidly.

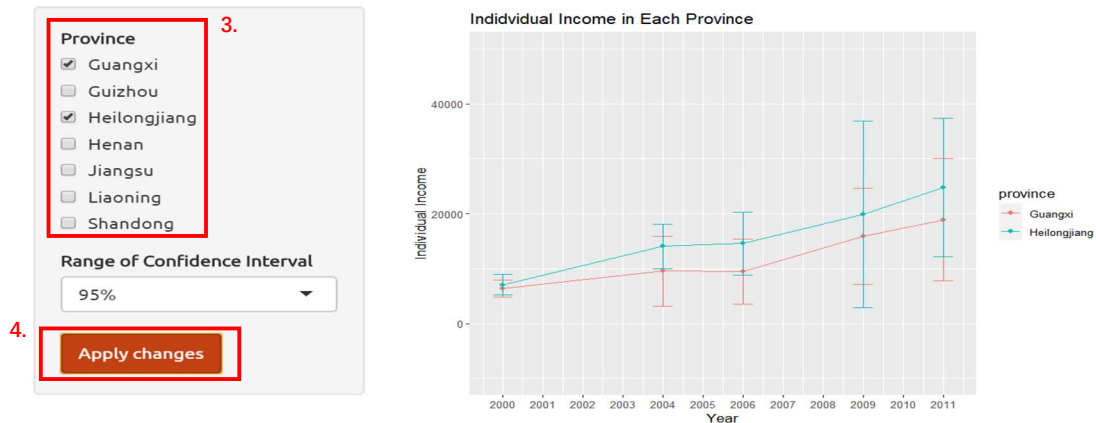


Figure 4 Individual Income comparing panel-2

We also use interactive on these charts. When you move mouse on lines or points of our charts, annotations will appear to illustrate some numbers (square 6). Users can select any province you are interested in and put in certain years you want to compare (square 5).

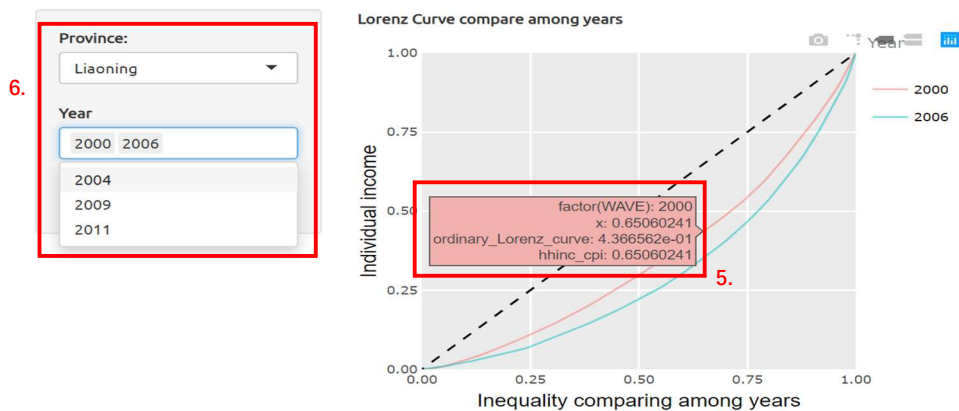


Figure 5 Gini index comparing panel-by years

Another interesting way is to fix result at certain year and choose different provinces you want to compare (square7).

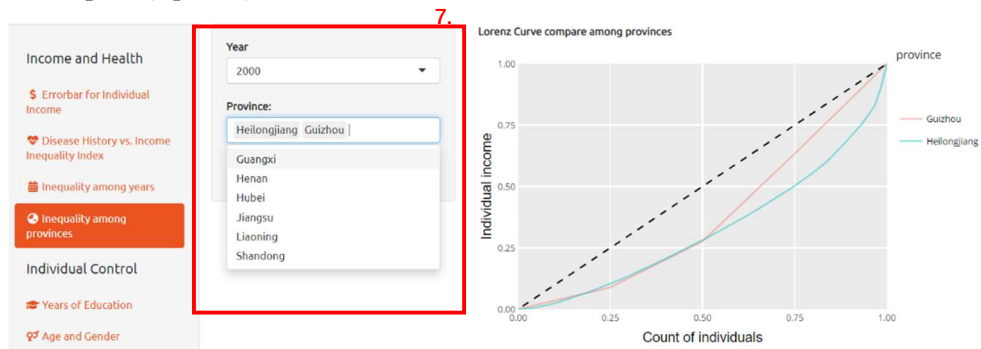


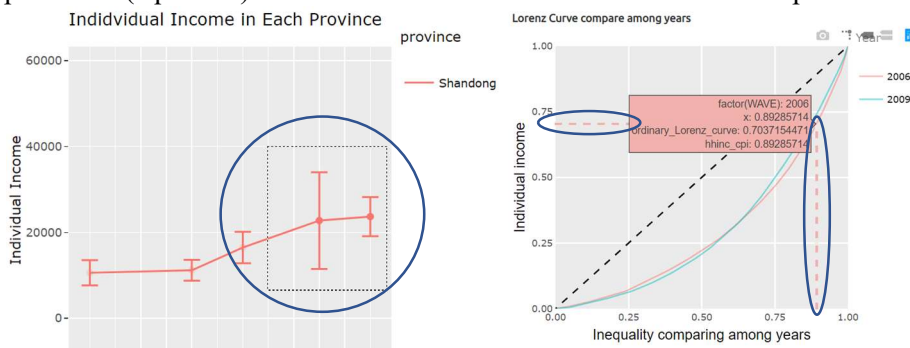
Figure 6 Gini index comparing panel-by provinces

When you focus on the right top corner of Error Bar chart or Lorenz Curve, you can see following image. This tool bar gives you more interactive functions which can do on our charts.



Figure 7 Control Panel of interactive plot

For example, we can use this photo symbol(square8) to save your plot to picture as .png format. Select tools (square9) to select certain area on charts and look into its detail. And use toggle spike lines (square10) to see clearer number on two axes for certain point.



We can also use this comparing button (square 11) to show all annotations on different lines and compared all the other at the same time.

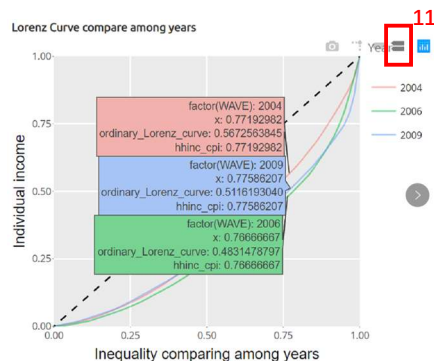


Figure 8 Mutual comparison pane

### 3 Pairwise correlation

This panel can help the user achieve two tasks: one task is to do pairwise correlation analysis; the other task is to work as preparation procedure for model building. The shades of color will present the correlated extent. The denser the color, the higher correlated of the two variables.

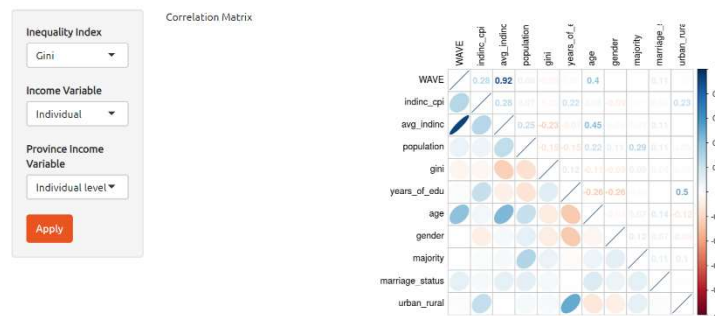


Figure 9 Pairwise correlation matrix

### 4 Panel Data Modelling

There are two sub-tab in ‘Panel Data Modelling’: Hypothesis test and Modelling.

#### 4.1 Hypothesis test

In this part, we offer much flexibility for researchers to do hypothesis test. This hypothesis test is conducted between continuous independent variables to the binary variable disease history. Default setting of DHIIVAT is 95% confidence interval. Beyond graphic representation by violin-box plot to show the distribution, our panel offers statistic calculation result.

In below example, we test ‘Age’ vs ‘Disease history’ using ‘parametric’ test method. Using t-test method,  $p\text{-value} < 0.001$ , we can reject null hypothesis. The null hypothesis here is : Age is not statistically significant to Disease history.

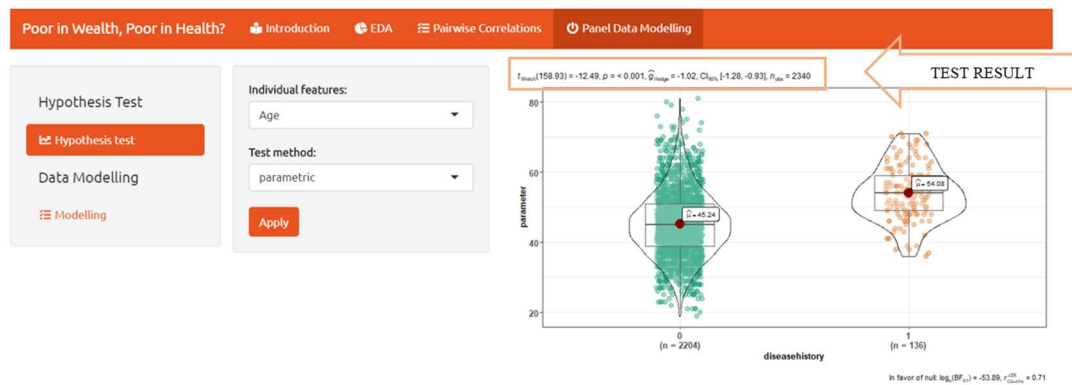


Figure 10 Hypothesis test pane

#### 4.2 Modelling

After analyzing all variables intercorrelation, the user has the understanding of selecting variables. Then we can start to build model in ‘modelling’ tab.

In following picture, you can follow the steps to build your own model. Steps are marked by heptagon:

Step 1: select independent inequity variables by drop-down menu

Step 2: Use checkbox to tick Individual Control Variables you want to include to your regression model and click ‘**Apply changes**’ button to run model result.

Step 3: You can see all model result table on the right side of the panel. Mark no.3 is to clarify ‘Fixed effect’, in DHIIVAT, the fixed effect is IDind and WAVE.

Step 4: This is Pseudo R<sup>2</sup> is to show the regression result, higher R<sup>2</sup> means better statistic result.

Step 5: In this toolkit, variables with \* means significant, the level of significance is stand for significance level. You can check for significant variables by referring to this symbol.

The screenshot displays the 'Model building pane' interface. On the left, there are three main sections: 'Hypothesis Test' (with a 'Hypothesis test' link), 'Data Modelling' (with a 'Modelling' button), and 'Hypothesis Test' (with a 'Hypothesis test' link). The central panel is titled 'Dependent Variable:' and has 'Disease History' selected. Below this are four categories of independent variables: 'Independent Inequity Variables:' (with 'Theil L Index' selected, marked with a '1'), 'Independent Income Variables:' (with 'Individual Income' selected), 'Independent Average Income Variables:' (with 'Average Individual Income' selected), and 'Individual Control Variables:' (with 'Years of Education', 'Age', 'Marriage Status', and 'Urban or Rural Residence' checked, and 'Occupation' and 'Occupation Type' unchecked, marked with a '2'). An 'Apply changes' button is at the bottom of this section. On the right, a regression results table is shown for the dependent variable 'diseasehistory'. The table lists coefficients and standard errors for various independent variables. The 'age' variable is highlighted with a '5' and has a coefficient of 0.078\*\*\*. The 'Fixed effects' section shows 'IDind, WAVE' with a '3'. The 'Pseudo R<sup>2</sup>' is 0.482, highlighted with a '4'. A note at the bottom indicates significance levels: \*p<0.1; \*\*p<0.05; \*\*\*p<0.01.

Dependent variable:	
diseasehistory	
theil_l	3.362 (3.552)
indinc_cpi	0.00000 (0.00000)
avg_indinc	-0.00002 (0.00002)
years_of_edu	0.030 (0.038)
age	0.078*** (0.017)
marriage_status	19.045 (5,461.659)
urban_rural	1.360*** (0.521)
Estimator	logit
Probability estimated	1
Fixed effects	IDind, WAVE
Std. errors clustered	No
Pseudo R <sup>2</sup>	0.482
Note:	*p<0.1; **p<0.05; ***p<0.01

Figure 11 Model building pane