

Instructions for TIDAL

Introduction

The aim is for this digital tool to facilitate trajectories work and remove barriers to implementing longitudinal research to researchers without specialist statistical backgrounds. It helps guide trajectory modelling and capture clinically meaningful features from mental health trajectories for specific individuals and/or specific groups of people.

Installation and usage

Locally

Please install the R package and launch the Shiny app locally if you want to upload sensitive data. If using R Studio it's recommended to restart your R session before installing.

```
# install.packages("remotes")
remotes::install_github("TIDAL-modelling/TIDAL")
# Note if prompted to update packages you can select option 3/None.
# Updating all packages (option 1) might take a while to run.
library("TIDAL")
# Launch the R Shiny app
launchTIDAL()
# To get documentation for launchTIDAL()
?launchTIDAL
```

Online

<https://tidal.shinyapps.io/tidalapp/>

To use this tool online please do not upload any sensitive data. Only use the synthetic datasets, described below.

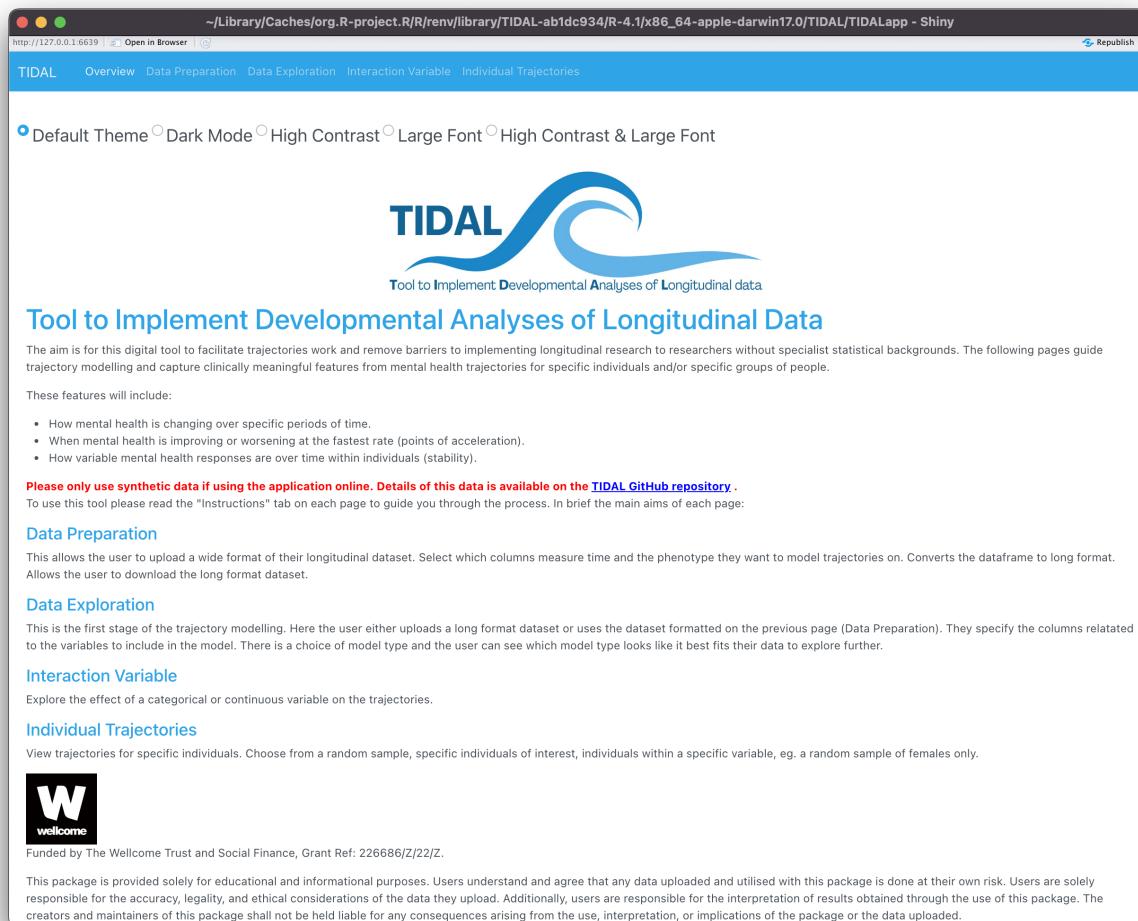
Synthetic datasets

A synthetic dataset is available called “emot_reg_emot_simulated.csv” this is also embedded within the app. Details about this synthetic dataset can be found on GitHub in the /data subfolder here.

Interacting with the app

Overview page

When you launch the app the first page of the app looks like this:



Please read through this all this information before preceding with the following pages.

Data Preparation page

- Click on the Data Preparation tab. This page is for converting data from a wide format into a long format, which is typically needed for longitudinal analysis. If you already have data in long format then proceed to Data Exploration. However, if you have wide format data or would like to explore the synthetic dataset we have provided embedded with in the app then please continue below.
- Firstly, please read the instructions in the Instructions tab on this page. Either upload your own data (in .csv or .tsv format) or click on “Use demo synthetic data”.
- In this demo we will use the synthetic data.

~/Library/Caches/org.R-project.R/R/renv/library/TIDAL-ab1dc934/R-4.1/x86_64-apple-darwin17.0/TIDAL/TIDALapp - Shiny

http://127.0.0.1:6639 | Open in Browser | Republish

TIDAL Overview Data Preparation Data Exploration Interaction Variable Individual Trajectories

Select data source:
 Upload data
 Use demo synthetic data

Select column for participant ID:

Select columns for age at each time point:

Select columns for the variable to model trajectories on at each time point eg. depression scores:

Name of new column for age:

Name of new column for time point:

Name of new column for variable to model trajectories on:

Impute missing age ?

Instructions Output

Convert longitudinal data from wide format to long format

Please only use synthetic data if using the application online. Details of this data is available on the [TIDAL GitHub repository](#).

In order to model trajectories R requires the data frame to be in "long" format. It's likely that the data is initially inputted as a "wide" format. This page allows you to upload wide formatted data and converts it to long. Similar to the image displayed below.

Wide

Participant ID	Age (time point 1)	Age (time point 2)	Age (time point 3)	Depression Score (time point 1)	Depression Score (time point 2)	Depression Score (time point 3)
A						
B						
C						

Long

Participant ID	Time Point	Age	Depression Score
A	1		
A	2		
A	3		
B	1		
B	2		
B	3		
C	1		
C	2		
C	3		

Steps:

1. Upload a wide format comma separated *.csv or tab delimited *txt or *tsv file of your dataset. It is recommended your longitudinal dataset has at least 4 time points. This must be less than 30 MB, column names must not have any spaces in them and missing data must be coded as "NA". Click on the "Output" tab and proceed with the following steps.
2. Select which columns correspond to age and the variable you want to model, eg. depression, at each time point. Make sure that you select these in the correct chronological order and have the same corresponding time points for the variable you want to model and age.
3. Default names for the new columns are "age", "time_point" and "score". You can change these in the text boxes provided if you wish, but make sure they are unique and also are not names of columns that already exist in your dataset.
4. If you have missing data for age there is an option to impute the mean from each time point for this variable. There is a tick box you can check to do this.
5. Now you can see a preview of the newly formatted long dataframe ("Output" tab). You have the option to download it (in .csv format) and also use for analysis on the subsequent pages.

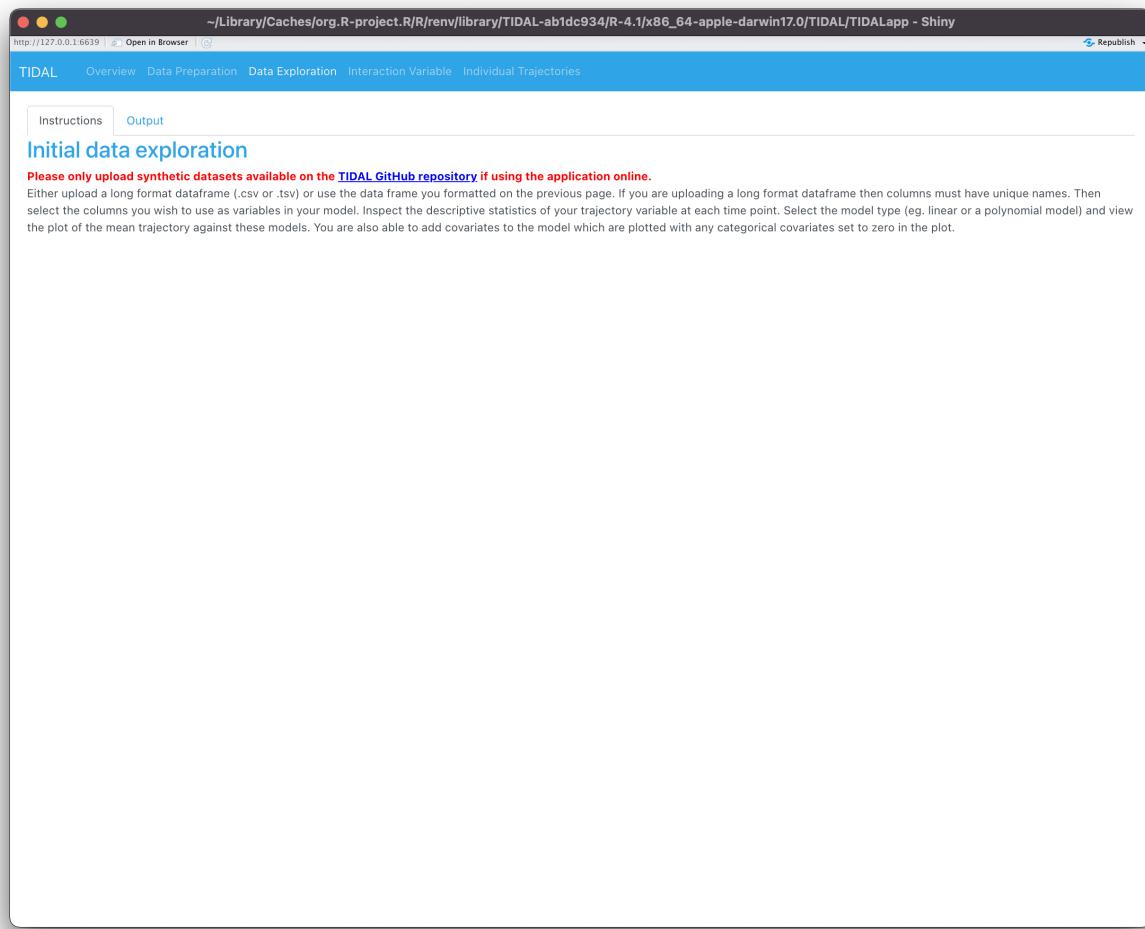
- Select columns for age at each time point. In the synthetic data sets we have 5 time points called "age_t1", "age_t2", "age_t3", "age_t4" and "age_t5". This is the age of the individual at each time point/assessment. Select these columns from the drop down menu as shown in the figure below.
- Then select columns for the questionnaire responses at each time point. In the synthetic data sets we have responses at 5 time points for the Strengths and Difficulties Questionnaire (SDQ). These 5 time points are called "sdq_t1", "sdq_t2", "sdq_t3", "sdq_t4", "sdq_t5". Select these columns from the drop down menu as shown in the figure below.
- There are the options to change the column names for age, time point and the new column made for the variable to model trajectories on.
- There is also the option to impute missing age with the mean age.
- Click on the Output tab to view a preview of the long format dataset, as shown in the figure below. Any warning messages will also be displayed in this Output tab.
- A download button will appear when you have the option to download this data as a .csv in long format. Otherwise you can continue to the next page.

The screenshot shows the TIDAL Data Preparation interface. On the left, there are several input fields and dropdown menus for selecting data sources, participant IDs, age columns, and trajectory variables. On the right, there are tabs for 'Instructions' and 'Output'. The 'Output' tab is active, displaying a preview of the first few rows of a long-formatted dataset. The columns shown are subject, time_point, age, and score. The data includes rows for various subjects (e.g., 2, 3, 17589, 17590) at different time points (e.g., age_t1 to age_t5), with corresponding age values and scores.

subject	time_point	age	score
2	age_t1	3.041	0
2	age_t2	5.482	2
2	age_t3	NA	NA
2	age_t4	11	0
2	age_t5	14	0
3	age_t1	3.126	2
...
17589	age_t5	NA	NA
17590	age_t1	3.107	4
17590	age_t2	5.142	2
17590	age_t3	6.989	1
17590	age_t4	10	2
17590	age_t5	14	4

Data Exploration page

- Click on Data Exploration tab and read the Instructions. Then click on the Output tab.

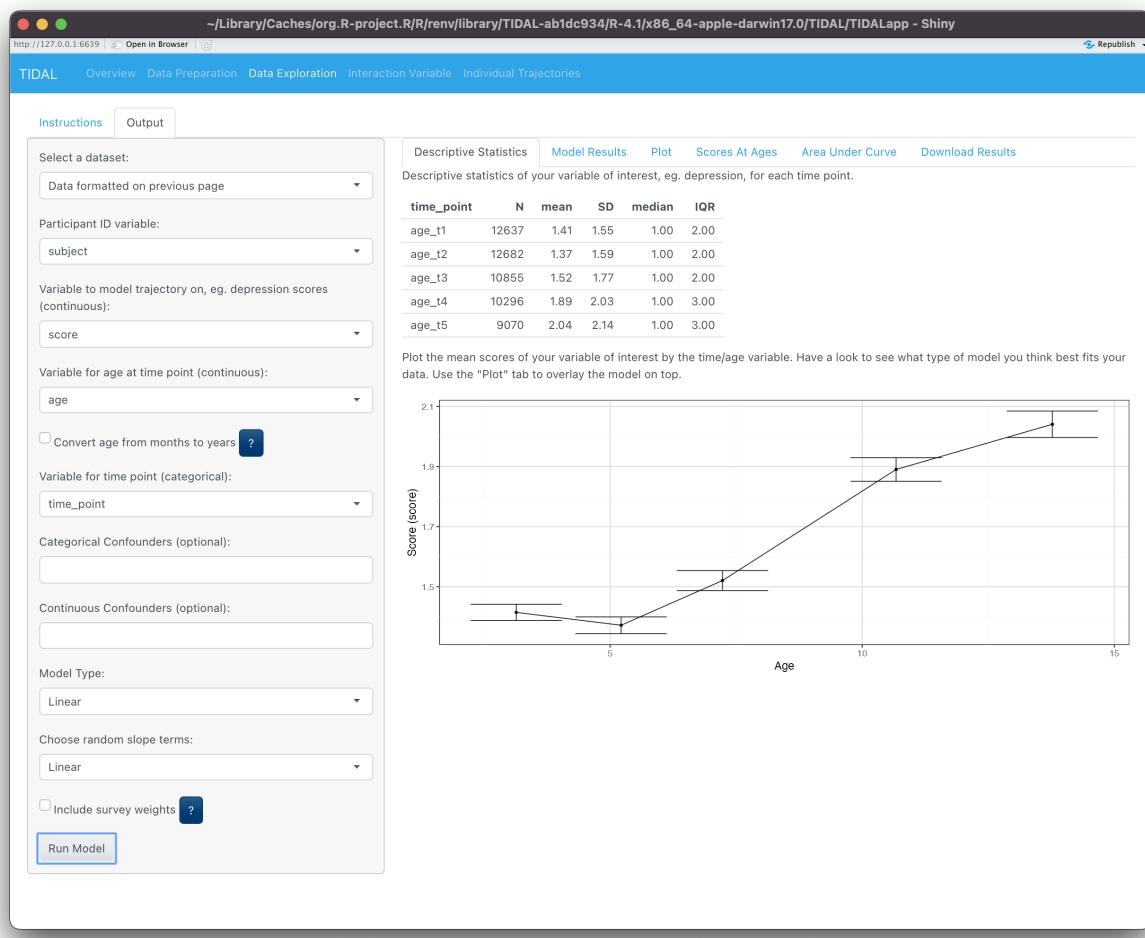


* Either upload a long format dataset (which you saved as a .csv from the previous page) or select “Data formatted on the previous page” on the drop down menu.

- In this example we will use the “Data formatted on the previous page”

The screenshot shows a Shiny application window titled "TIDAL" with the URL "http://127.0.0.1:6639". The top navigation bar includes links for "Overview", "Data Preparation", "Data Exploration", "Interaction Variable", and "Individual Trajectories". Below this, there are two tabs: "Instructions" (selected) and "Output". A sidebar on the left is titled "Select a dataset:" and contains three options: "Upload a long format dataset" (highlighted with a blue border), "Upload a long format dataset", and "Data formatted on previous page" (which is blue). To the right of the sidebar, a horizontal menu bar offers "Descriptive Statistics", "Model Results", "Plot", "Scores At Ages", "Area Under Curve", and "Download Results". Below the menu, descriptive text explains the purpose of each tab. The main area of the screen is currently empty.

- Select the correct columns from your dataset for “Participant ID”, “Variable to model trajectory on”, “Variable for age at time point”, “Model Type” - either Linear, Quadratic, Cubic or Quartic and “random slope terms”. There’s also the option to convert age from months to years, add covariates or include survey weights. When you have selected these choiced click “Run Model”.
- The Descriptive Statistics tab below shows some summary statistics about your data and a plot.



- Model Results tab returns the the code that was run, the fixed and random effects and some text for interpreting those results.

~/Library/Caches/org.R-project.R/R/renv/library/TIDAL-ab1dc934/R-4.1/x86_64-apple-darwin17.0/TIDAL/TIDALapp - Shiny

<http://127.0.0.1:6639> | Open in Browser | Republish

TIDAL Overview Data Preparation Data Exploration Interaction Variable Individual Trajectories

Instructions Output

Select a dataset:
Data formatted on previous page

Participant ID variable:
subject

Variable to model trajectory on, eg. depression scores (continuous):
score

Variable for age at time point (continuous):
age

Convert age from months to years ?

Variable for time point (categorical):
time_point

Categorical Confounders (optional):

Continuous Confounders (optional):

Model Type:
Linear

Choose random slope terms:
Linear

Include survey weights ?

Run Model

Descriptive Statistics Model Results Plot Scores At Ages Area Under Curve Download Results

Model Formula: score ~ age + (1 + age | subject)

```
lmer(formula = score ~ age + (1 + age|subject),
      REML = FALSE,
      data = newModelData,
      control = lmerControl(optimizer="bobyqa",
                            optCtrl=list(maxfun=2e5)))
```

Please see more information about the "bobyqa" optimiser [here](#). The use of alternative optimisers is not currently supported. The argument `REML = FALSE` indicates the model was fitted by maximum likelihood.

Number of observations and groups
The number of observations (measurements) is 55,537 and the number of groups (people) is 12,720.

Fixed Effects

effect	term	estimate	std.error	statistic	2.5 %	97.5 %	p.z
fixed	(Intercept)	1.618	0.012	135.279	1.595	1.642	p < 0.001
fixed	age	0.066	0.002	31.950	0.062	0.070	p < 0.001

The score at the intercept is 1.62. The intercept here has been shifted to the mean age of all the assessments which is 7.58. You could interpret this as the score at the intercept of age 7.58 is 1.62.

Every unit increase in age is associated with an increase of score by 0.07.

The model fit (deviance) is 208455.06, you can compare this value to other similar models to determine which model has a better fit.

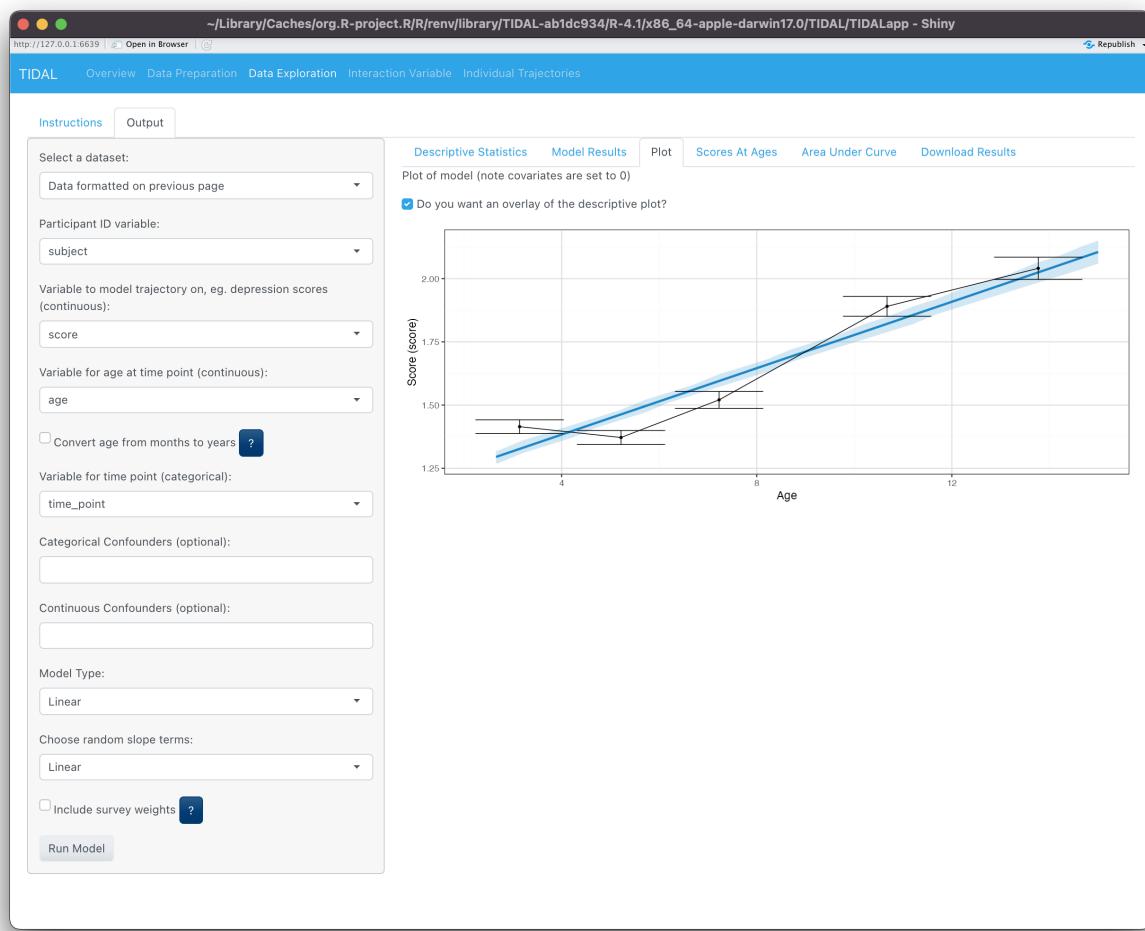
Random Effects

Level	Variable1	Variable2	Variance/Covariance	SD Variance/Covariance
subject	(Intercept)	NA	1.404	1.185
subject	(Intercept)	age	0.083	0.468
subject	age	NA	0.022	0.149
Residual	NA	NA	1.563	1.250

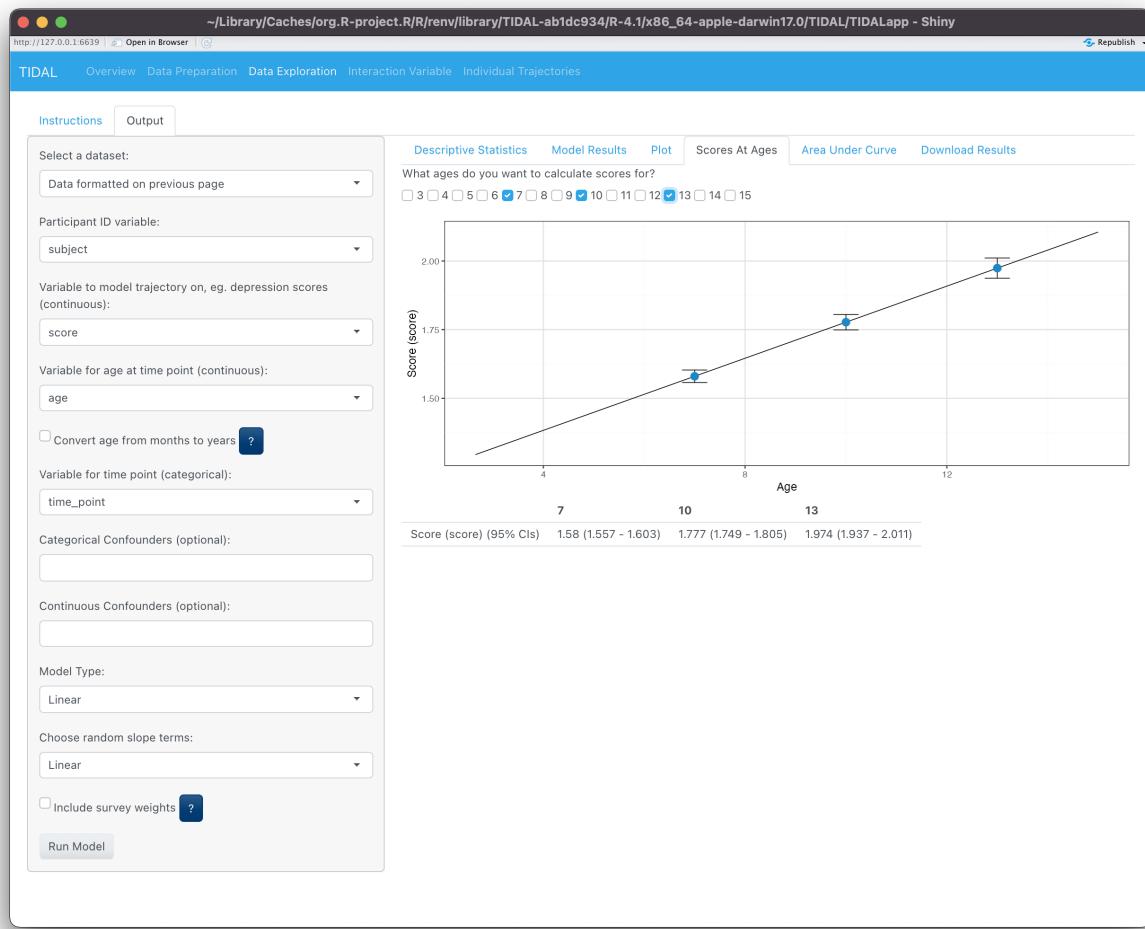
The intercept variance (how much variability there is between individuals for their intercepts) for your model is 1.404. The covariance between the intercept and age is 0.083. The age variance (how much variability there is between individuals for their age) is 0.022.

The residual variance (how much variability there is within individuals) from your model is 1.563.

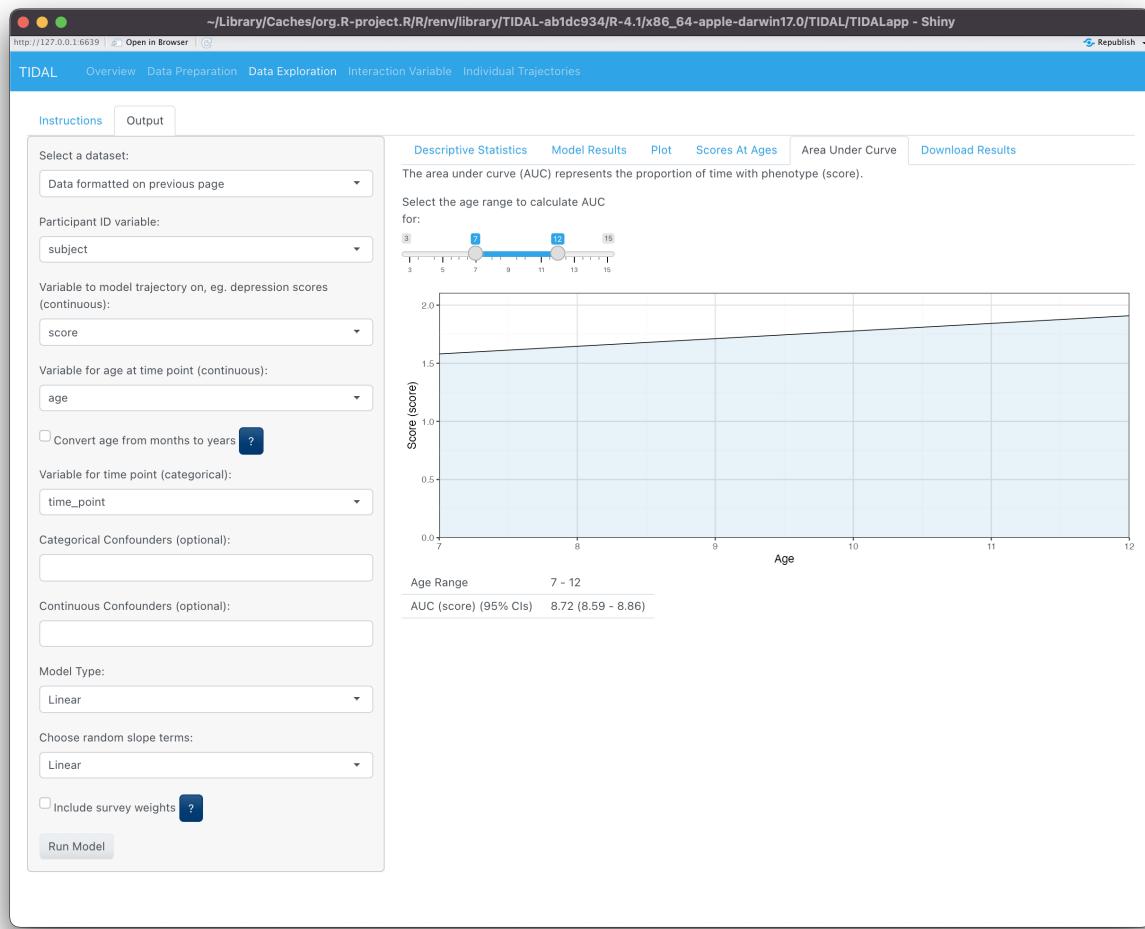
- The Plot tab shows the model trajectory plotted with the option of overlaying the plot of descriptive data.



- The Scores At Ages tab allows you to explore the trajectory in more ways. By showing the calculated estimates of the questionnaire score for a given age. Click on the check boxes to select which ages you want to calculate scores for.



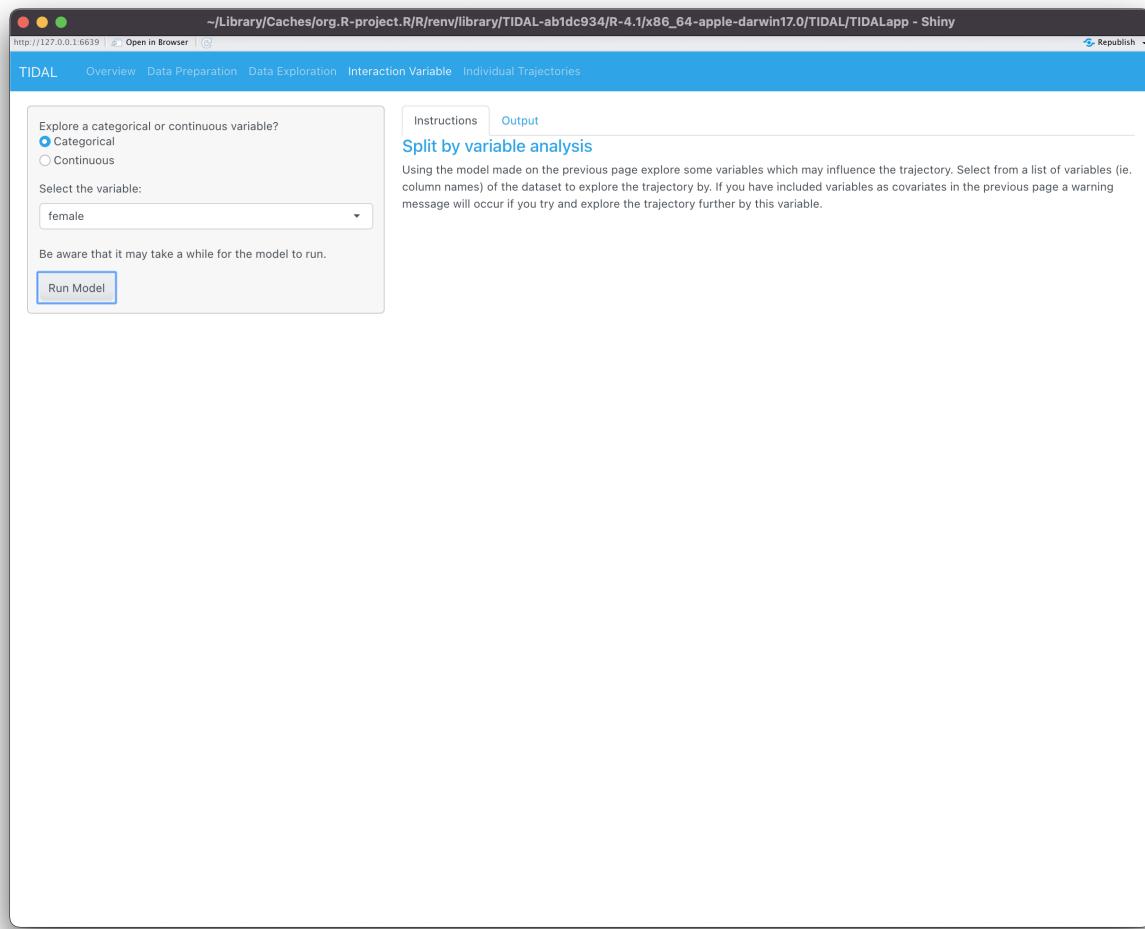
- The Area Under Curve (AUC) tab allows you to look at the proportion of time spent with that trait you are modelling. There is a slider bar to select which ages you want to calculate AUC for and the calculated estimates of this is displayed in a table below the plot.



- Finally, you can click on Download Results if you wish to download a pdf report of this. Note that you may have to have LaTeX installed if running this locally.

Interaction Variable page

- When you have your base model and have completed the Data Exploration you can click on the Interaction Variable tab. Using the model made on the previous page explore some variables which may influence the trajectory. Select from a list of variables (ie. column names) of the dataset to explore the trajectory by. If you have included variables as covariates in the previous page a warning message will occur if you try and explore the trajectory further by this variable.
- In the example below we will walk through exploring a Categorical variable to split by, but the same approach can be applied with a Continuous variable.



- When you have selected a variable click Run Model and view the results in the Output tab. Similar to the Data Exploration page there are tabs to interact with in the same way including: Plot, Scores At Age, Area Under Curve and Download Results, as seen in the figures below.

~/Library/Caches/org.R-project.R/R/renv/library/TIDAL-ab1dc934/R-4.1/x86_64-apple-darwin17.0/TIDAL/TIDALapp - Shiny

<http://127.0.0.1:6639> | [Open in Browser](#) | [Help](#)

[Republish](#)

TIDAL Overview Data Preparation Data Exploration Interaction Variable Individual Trajectories

Explore a categorical or continuous variable?

Categorical
 Continuous

Select the variable:

female

Be aware that it may take a while for the model to run.

[Run Model](#)

Instructions Output

Model Results Plot Scores At Ages Area Under Curve Download Results

Model Formula: score ~ age + (1 + age | subject) + female + age * female

Fixed Effects

effect	term	estimate	std.error	statistic	2.5 %	97.5 %	p.z
fixed	(Intercept)	1.527	0.017	90.961	1.494	1.560	p < 0.001
fixed	age	0.047	0.003	16.441	0.042	0.053	p < 0.001
fixed	female1	0.184	0.024	7.729	0.138	0.231	p < 0.001
fixed	age:female1	0.036	0.004	8.911	0.028	0.044	p < 0.001

The interaction variable you have chosen has been factorised with the lowest level "female0" being the reference or baseline category. For "female0", the score at the intercept is 1.53. The intercept here has been shifted to the mean age of all the assessments which is 7.58. You could interpret this as the score at the intercept for "female0" at 7.58 is 1.53.

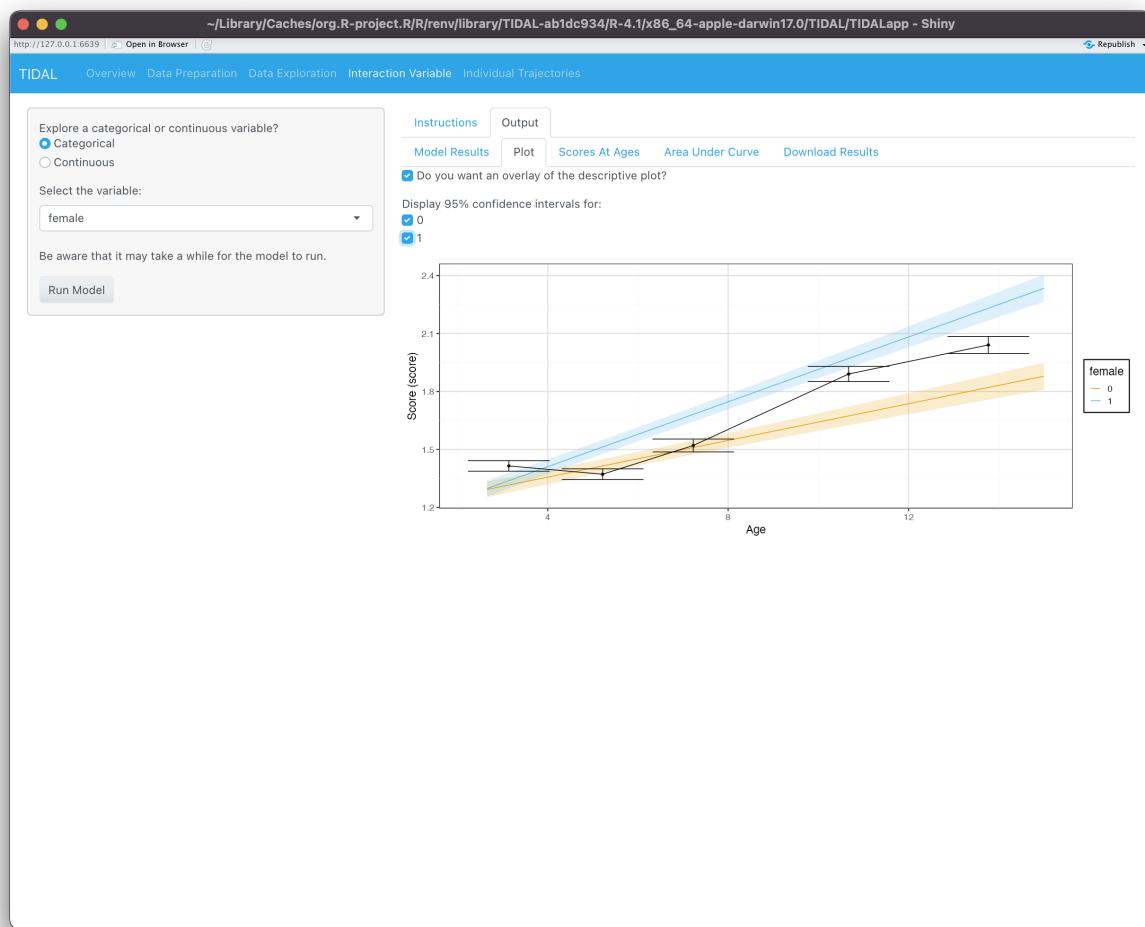
For "female0", every unit increase in age is associated with an increase of score by 0.05.

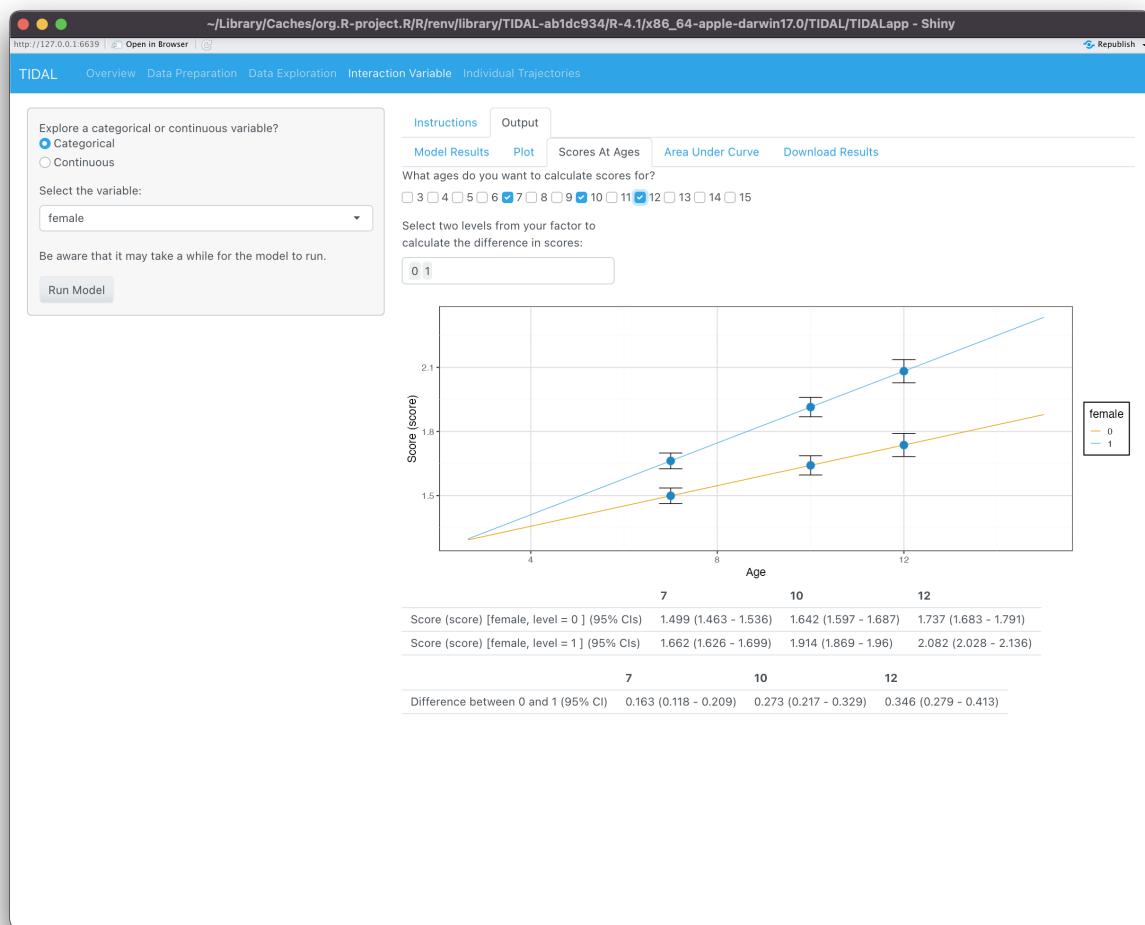
To estimate the effect of different trajectories, you can add the intercept and age estimates to the corresponding interactions and age:interactions to get group specific trajectories.

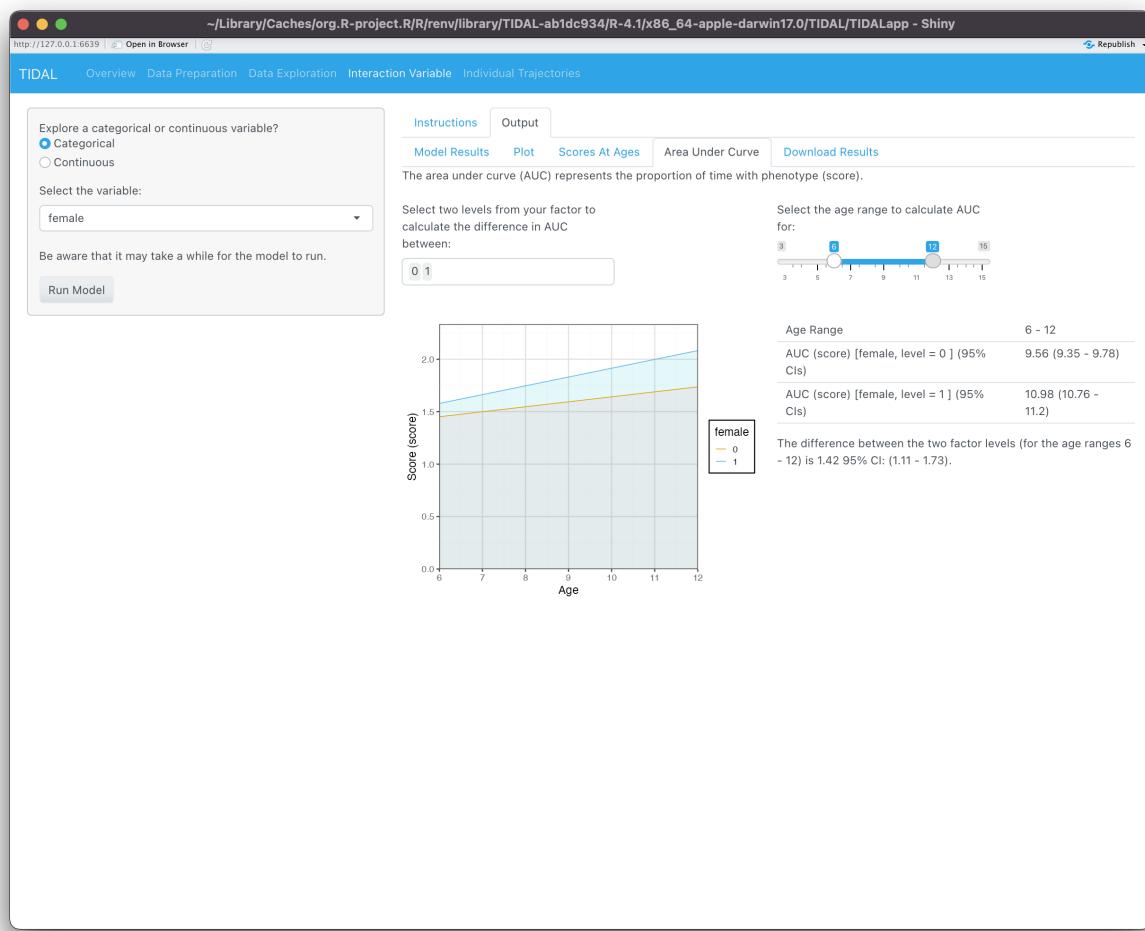
Further information on how to interpret these results can be found on the TIDAL GitHub training videos section. Please also see the "Plot" tab for visualisation of these results.

Random Effects

Level	Variable1	Variable2	Variance/Covariance	SD Variance/Covariance
subject	(Intercept)	NA	1.394	1.181
subject	(Intercept)	age	0.081	0.464
subject	age	NA	0.022	0.148
Residual	NA	NA	1.563	1.250

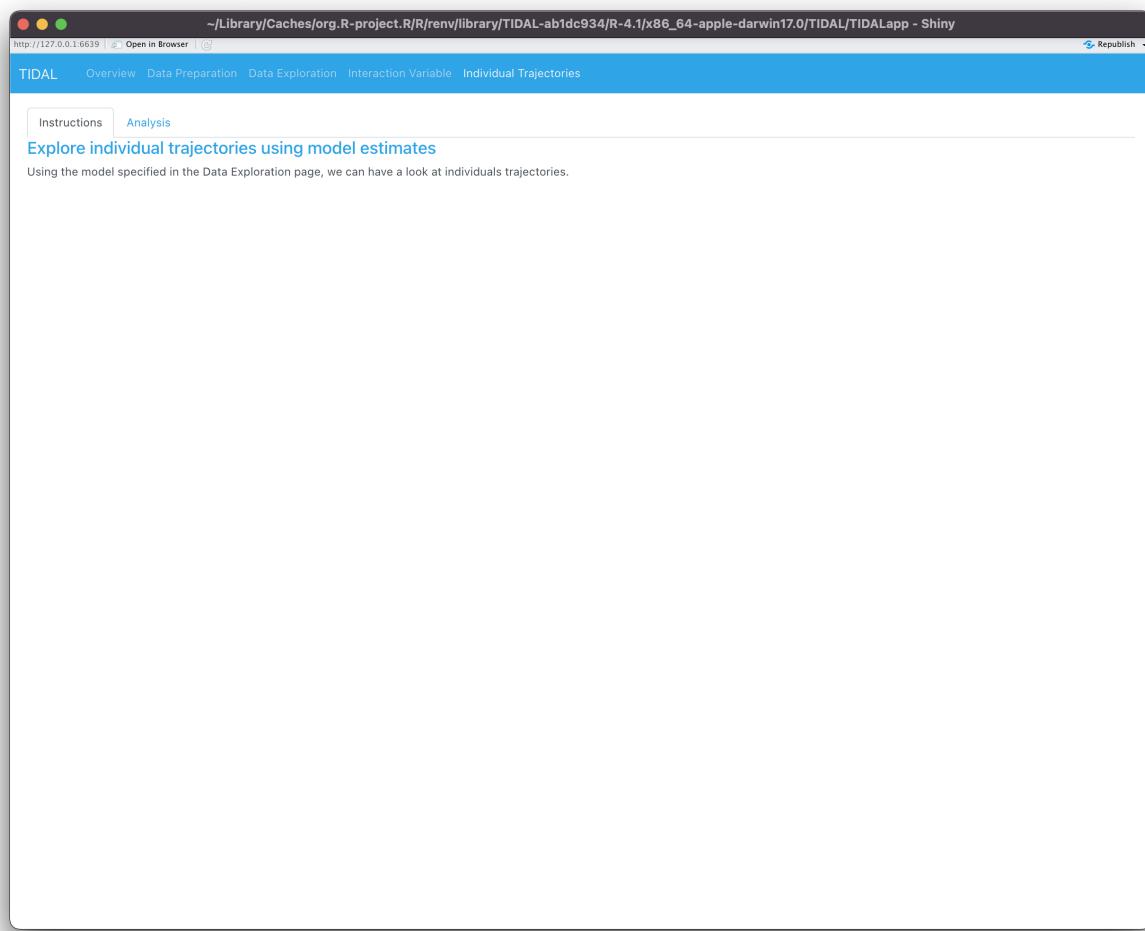




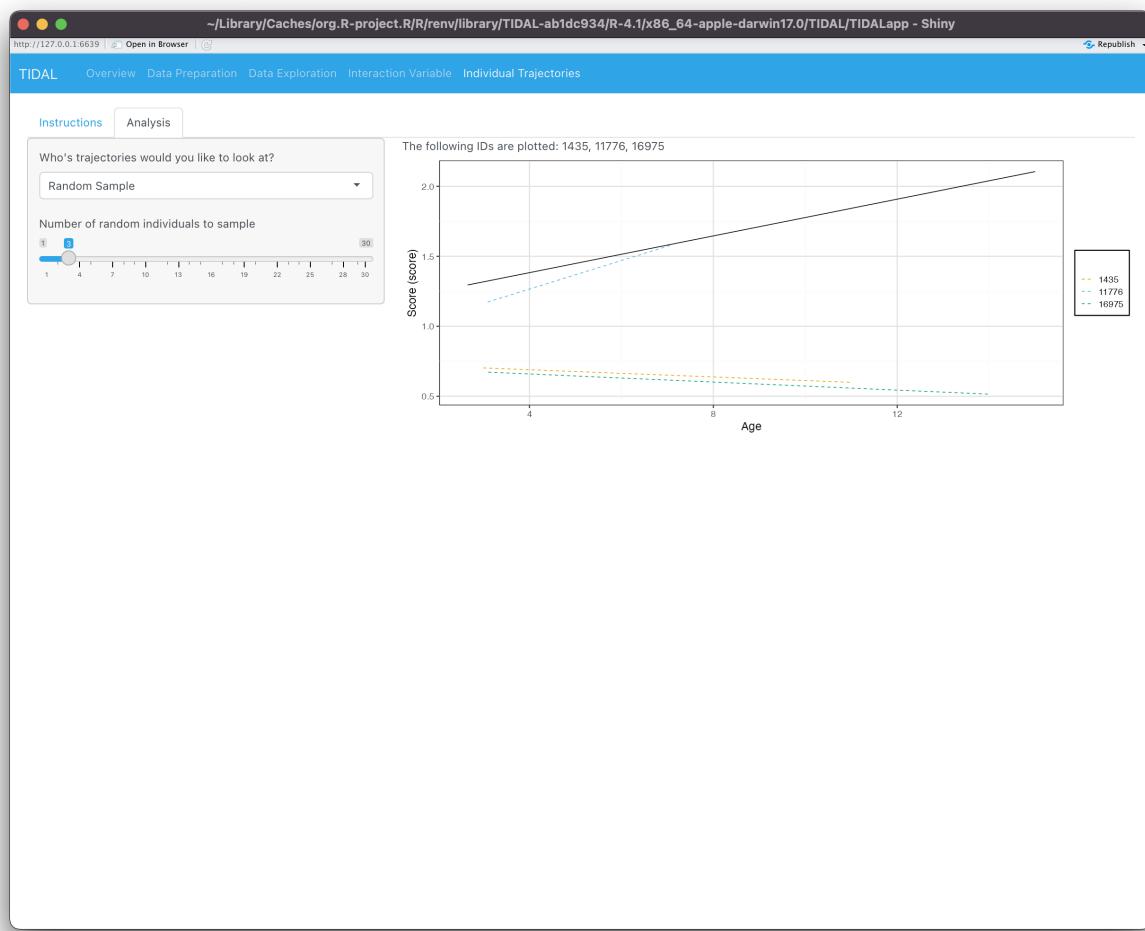


Individual Trajectories page

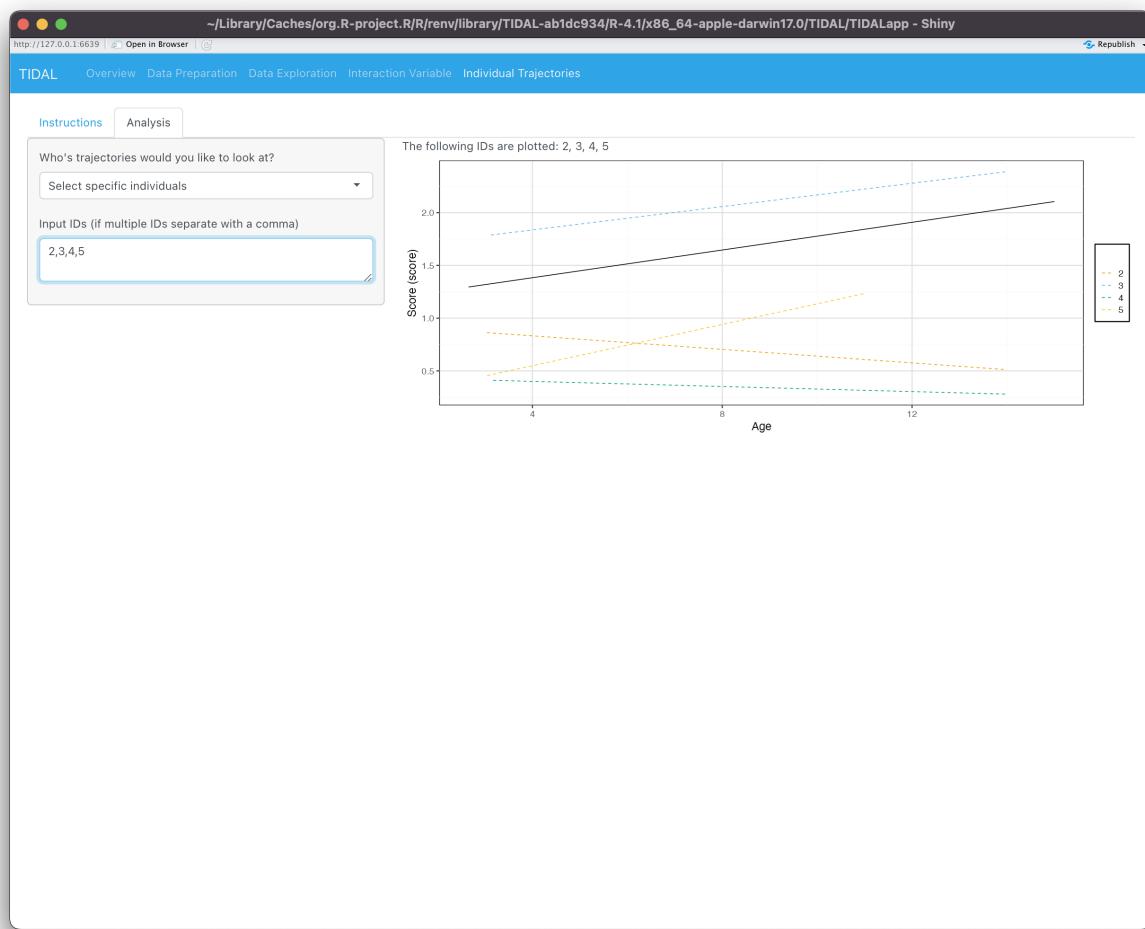
- Using the model specified in the Data Exploration page, we can have a look at individuals trajectories. Please click on the Analysis tab to look at this.



- From the drop down option box you can either look at a Random Sample of individuals and select the number of random individuals from the slider.



- Or you can select a specific set of individuals if you know their IDs. This will be a value in the column you specified as the Participant ID.



- Or you can look at A Specific Variable and choose a variable of interest and the level from that variable.

