

Shiny App: DATA2902 Survey Visualisations and Hypothesis Tests

App: <https://becowens.shinyapps.io/data1/>

Git: <https://github.com/becowens/data1>

The Shiny app allows users to explore survey data interactively by selecting any two variables from the dataset. It provides detailed summaries for both numerical and categorical variables - including descriptive statistics, levels, and cleaning information. Users can visualise data through bar charts, box plots, or scatter plots depending on variable types, with the option to filter which categorical levels to display. Statistical testing is included: t-tests for numeric vs categorical variables (with two levels) and chi-square tests for categorical vs categorical variables, while numeric vs numeric pairs display the R^2 correlation. Graph titles and legends use short variable names for clarity, and colour coding helps distinguish variable types and categories. To reduce clutter and improve readability, only the top 10 categories for categorical variables are displayed in graphs. Numeric variables and categorical variables with many levels cleaned prior to the app running: outliers are removed using z-scores, missing values are dropped and categorical levels are ordered to ensure logical sequences.

Limitations include the restriction of t-tests to only two levels of categorical variables and the lack of hypothesis testing for numeric-numeric pairs, which only show correlation. Graphing is limited to basic plot types and does not contain advanced interaction features like hover, zoom or lasso.

Choose Variable 1

Choose Variable 2

Significance level (alpha)

Run Tests

☐ Show warnings

Variable Information

Graph

Hypothesis Test

There are two key selection tabs - The first allows for the selection of variables and to customise your input.

The second allows you to choose how to view your selected variables - it includes the following tabs.

Variable Information : This pane gives a detailed summary for both variables, including variable names, descriptions, how the data was cleaned and how many data points were present before and after cleaning. For numerical variables, you'll see the full range, median, mean, min, and max. Categorical Variables will show all level options.

Graph: The data is represented visually. Box plots, bar charts or scatter plots appear according to your variable types. Colours match your category selections, and graphs update instantly as you select levels.

Hypothesis Test: This tab shows all your statistical test results. You'll see the output of t-tests or chi-square tests (depending on your variables) and a statement about whether your results are statistically significant. For numeric pairs the R^2 correlation value is shown

Choose Variable 1

Unit (Categorical)

Exercise Hours (Numeric)

Work Hours (Numeric)

Submit On Time (Categorical)

Used R Before (Categorical)

Team Role (Categorical)

University Year (Categorical)

Favourite Anime (Categorical)

Fluent Languages (Numeric)

Readable Languages (Numeric)

Run Tests

When you first open the app, you'll see dropdowns for "Choose Variable 1" and "Choose Variable 2", which let you select any variable in your dataset. These are colour-coded categorical variables are shown in blue and numeric ones in greens so you can always tell what data type you're working with. You must choose two different variables to get started; the app will remind you if you try to use the same one twice

Select levels to include for Used R Before

- ☒ No
- ☒ Yes

Select levels to include for University Year

- ☒ First year
- ☒ Second year
- ☒ Third year
- ☒ Fourth year
- ☒ Fifth or higher year

After selecting your variables, you'll see additional options appear if either is categorical. Here, you can include or exclude specific levels for each variable by just ticking or unticking the boxes. Only ticked variables will appear on the graph.

Significance level (alpha)

0.05

There is also a box to set your preferred significance level (alpha) beneath the variables. By default this is 0.05, just type a different number to change to a higher or lower value, or use the arrows to work in increments of 0.05. The hypothesis test will automatically determine if your test is significant at your selected alpha level.

Choose exactly two levels to compare:

First year

Second year

Third year

Fourth year

Fifth or higher year

If you're running a t-test using a categorical group with more than 2 levels, this pop-up prompts you to choose precisely two groups to compare. Use the dropdown to select the exact categories you want to test. The app won't let you continue until you've picked exactly two.

- ☒ Show warnings

Chi-square warning: Chi-squared

There's also a checkbox for "Show warnings", tick this if you want to see details about statistical test assumptions, like if some cell counts are too small for reliable chi-square results

Generative AI (ChatGPT) was used to assist with adding print statements for debugging, renaming long variable names to short forms, creating placeholder variable information for testing displays, and generating a blue colour palette for the graphs.