India TB Dataset: Workshop Instructions

Introduction

Welcome! For today's workshop, you'll create a short Quarto-based report exploring a dataset on Tuberculosis treatment pathways in India, incorporating data visualization with plotly.express and tables with itables.

Setup for Easy Instruction-Following

Download this instructions file to your computer, then add it to your graph_courses_python folder in VS Code. Then in the extensions panel, install the "vscode-pdf" extension. This should allow you open these instructions directly in VS Code, making it easier to follow along.

Steps

1) In your graph_courses_python folder, create a new Quarto file (.qmd) called `week_04_workshop.qmd` using the side panel.
2) Copy this YAML header into your file:

```
---
title: "India TB Pathways Analysis"
format: html
author: "YOUR NAME HERE"
---
```

(You may need to update this YAML later to change the execution options.)

3) Import the required packages (pandas, plotly.express, itables). (You should have installed these previously. But install them if you need to. For plotly.express, you install `plotly` not `plotly.express`.)
4) Under an appropriate section heading, add the following dataset description about the India TB Pathways dataset:

The India TB Pathways dataset is based on a study conducted in Vellore district, Tamil Nadu, India. It examines the pre-diagnostic costs and health-seeking behavior of 880 new adult pulmonary tuberculosis patients registered at Primary Health Centres (PHCs).

5) Download the dataset from this link, unzip it (extract all on Windows), then copy the file into a data folder in your graph_courses_python repository.

https://github.com/the-graph-courses/pbb_2025_q1_materials/raw/refs/heads/main/week_04_workshop/data/india_tb_pathways_and_costs.zip

(You can find out more about the dataset here if you're interested.)

6) Load the data using pd.read_csv()
7) Use the show() function from itables to display the first few rows of your dataset
8) Download the following image of the Tamil Nadu map and add it to an images folder in your repository:

https://github.com/the-graph-courses/pbb_2025_q1_materials/blob/main/week_04_workshop/images/tndistrict_1.png

9) Embed that image in your report using the appropriate Quarto syntax. Include a relevant description.

Here is an example of how to embed an image:

```
![Map of Tamil Nadu. Vellore district is in the north-east.](images/PATH_TO_YOUR_IMAGE_FILE)
```

10) Under an appropriate section heading, create a bivariate plot using Plotly Express that shows a relationship between two variables of your choice in the dataset. Add a brief interpretation of what your plot reveals.
11) The most important part: Render your report to an HTML format. Make sure it renders successfully. Your code should not be visible in the rendered output.

OPTIONAL CHALLENGE: Highlight Key Numbers with Inline Code

Use inline Python code to enhance your narrative. You might choose to highlight, for example, the maximum and minimum values of your numerical variables, or the most common values of your categorical variables.

Reference the Quarto documentation for guidance: https://quarto.org/docs/computations/inline-code.html

Here's a quick example using the tips dataset to show how to extract useful information:

```
tips = px.data.tips()
max_tip = int(tips.tip.max()) # int needed to print cleanly
common_day = tips.day.mode()[0]
```

You can then use these variables in your narrative like this:

```
The max tip amount was `{python} max_tip`.

The most common day in the tips dataset was `{python} common_day`.
```

The values there will then be dynamically inserted into your narrative. This allows you to highlight key numbers in your report without having to manually update the text every time you re-run your code or your dataset changes.