**RA Task**

**Data Generation**

We want to use this source of data, ERA5 monthly averaged data on pressure levels from 2023, to generate the downup variable for Punjab, India in 2023. This downup variable should be calculated based on the area of Punjab’s districts under wind influence. We want you to return a PDF, HTML, or Markdown document where you show your scripts and all the decisions you made. You should be very detailed in all your steps, argue all the decisions you take. This task uses large datasets, and we want to see how you manage them.

1.Obtain the data for U and V wind components for 2023 at the monthly level: U-component of wind, V-component of wind from this source: <https://cds.climate.copernicus.eu/datasets/reanalysis-era5-pressure-levels-monthly-means?tab=download>

2. Get the number of fires for Punjab India from <https://firms.modaps.eosdis.nasa.gov/download/>.

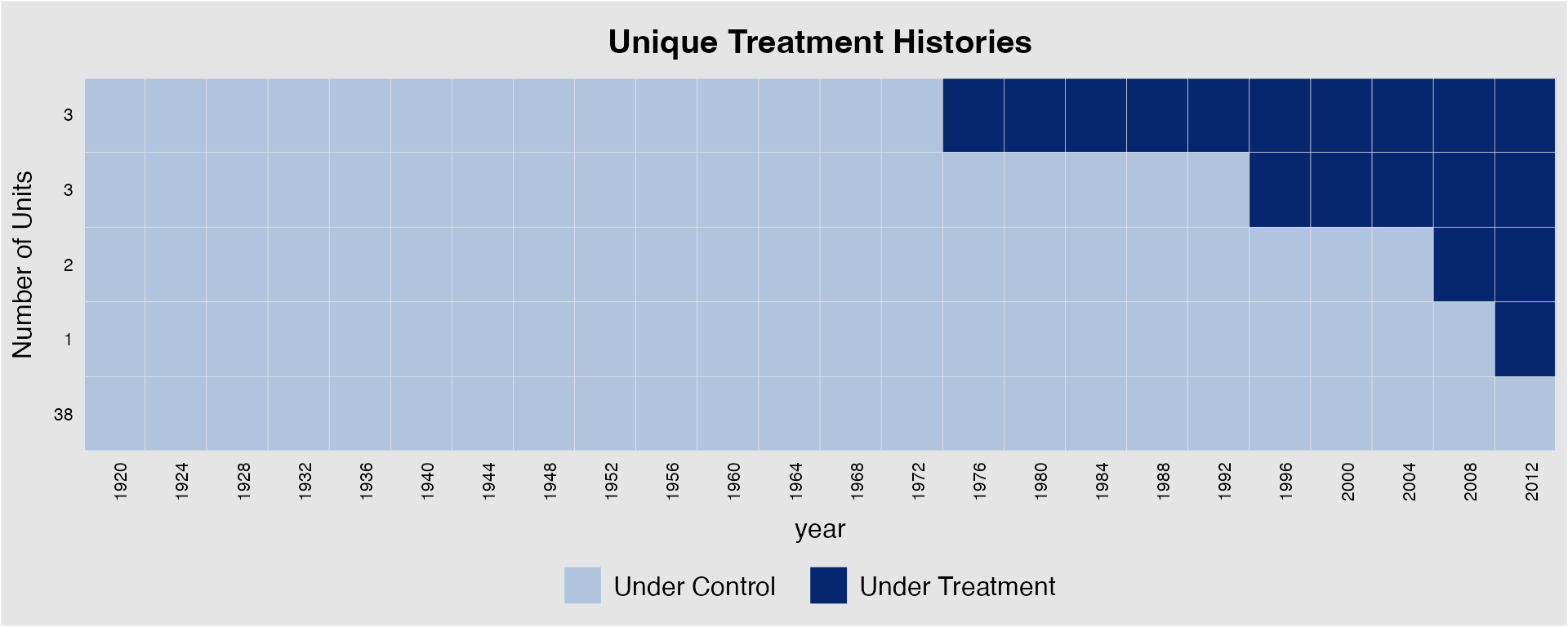
3. Split wind grids into 25 smaller grids. Keep only the grids that touch Punjab. Use the shapefile of India available at <https://www.devdatalab.org/shrug>.

4.Calculate the wind direction using the U and V wind components, and also the number of fires for each grid. Please, consider 0 when there is no fires.

5. Define the downup variable as 1 at the grid-month level if the grid pollutes most of the district area, and 0 otherwise. Using the grid centroid as the reference point, and based on the wind direction, determine whether most of each Punjab district’s area lies downwind. Return the calculated upwind and downwind areas. Set downup = 1 if the downwind area exceeds the upwind area. You should implement parallel computing for this step.

**Analysis**

1. Generate a plot like panelView but using xarray in Python or Raster in R. This is the opportunity for you to show your ability with these programing languages, and how you work with new technologies. We want to see the grids in the y axis and the months in the x axis. The cell should be color as gray when downup = 0, and blue when downup =1. We want a plot similar to the one below.



2. Difference-in-differences:(a) Write out the estimating equation for the difference-in-differences design to estimate the effect of being a downwind grid on the number of fires. Use the necessary subscripts and explain what the relevant terms mean.(b) Estimate the causal effect of polluting your own district on crop burning (measured by the number of fires). Downwind grid-cells are considered treated and, viceversa, upwindgrid-cells are considered controls. Produce a standard regression table with the results and the standard errors (clustered at the grid and month level). Remember that standard regression tables often include rows that tell the number of observations, the presence of fixed effects, the presence of control variables, the number of clusters, etc. (note that we don’t give you any control variables for this exercise).

3. Eventstudy Plot:(a) Run the relevant regression for an event study with a window of +/-3 months around the switch to treatment. Use the month of the switch as omitted category.(b) Using the result of the previous point, create a plot. Remember that event study plots will often have a dotted line at the omitted category and will also include 95% confidence intervals bars for the estimates.