Q2

#My code could scale up with larger datasets.

#First, instead of importing datasets one by one, I used for loop to automatically retrieve datasets from folders. For larger datasets, as long as they have keywords such as "rainfall" in their names, they can be imported via for loop.

#Second, I drop duplicate grid points before calculating their distances from district centers. This allows the computer to finish calculations in a much shorter time. For larger datasets, it is important to also drop those duplicates before computing the distances.

#Third, we can also use ArcGis to map large numbers of grid points to districts’ shapefile.

#For a dataset with 339 districts and 112 years, and a smaller grid, it will take more computing power and longer time to finish the calculation.

Q4

1. Mortalityit= α + βTemperatureit + εit
2. It means a 1 degree Celsius increase is associated with a 0.05 increase in the mortality rate. This is statistically significant because the t statistic is the coefficient divided by SE which is 2.5. Getting a t-value of 2.5 would be significant, as the alpha would be <0.05.
3. I can add a Temperature2 variable to the model. Or I can also add an age2 to the model if nonlinearities exist.
4. No. First, there are other confounding variables that might influence the relationship between temperature and mortality that are not included in the model, such as AQI(Air quality index). Second, if there is only temperature included in the model, we cannot conclude a causal relationship between temperature and mortality.
5. We can use TWFE model to remove potentially confounding factors. By adding time-fixed effect dummy variable, we can control for factors that remain unchanged relative to the year. By adding district-fixed effect dummy variable, we can control for factors that remain unchanged relative to the district. We can also add an interactive fixed effects variable which is the state\*year. This can control for unobservable factors that change over time at the state level.
6. We can also add variables such as AQI and age to remove potentially confounding factors. Because AQI and age can also influence mortality.
7. Mortalityit= α + βTemperatureit + δAQIit + ƬAgeit + ci + λt + ɵjt + εit

t = 1,2,3…T i = 1,2,3…N

ci , λt are fixed effects. ci is district fixed effect, λt is time fixed effect.

ɵjt is the interactive fixed effect. it is the state-level \* year to avoid multicollinearity.

εit is the error term.

The distribution of the data in different districts is different from each other, therefore, the standard error of each district is different. We need to consider heteroscedasticity. Therefore, we need to cluster the robust standard error to the district level.