Instructions: figure of shares of sales

Input file

1. "JPV_JIG\Trade\0-Raw_Data\Sales\sales_states.csv.xlsx". The file has 6 columns. The year, the identity of the firm that sells the product "seller", the identity of the firm that buys the product "buyer", the province where the seller lives "state_seller", the province where the buyer lives "state_buyer", and finally, the amount of the sale "amount".

Some notation:

Let's refer to firms (seller or buyer) with lower case letters and to provinces to upper case letters. Let's refer to firm i in province P as $i \in P$. Finally, the amount that firm $i \in P$ sells to firm $j \in P'$ would be $X_{ij,PP'}$ (the first subscript represents the seller, the second one the buyer. Then the third one the province of the seller, and the last one the province of the buyer). The unit of observation would be the combination state_buyer-state_seller. So, for example, PP' is one unit of observation. Since there are 20 provinces, then there are 400 units of observation (20 times 20).

For each unit of observation, compute:

1. The number of sellers in the unit of observation PP'. This is

$$n_{PP'} \equiv \sum_{j \in P'} \sum_{i \in P} \mathbb{I}(X_{ij,PP'} > 0),$$

   where $\mathbb{I}(X_{ij,PP'} > 0) = 1$ if $X_{ij,PP'} > 0$ and zero otherwise.

2. The share of the largest seller in PP'. This is

$$\chi_{PP'} \equiv max \left\{ \frac{\sum_{j \in P'} X_{ij,PP'}}{\sum_{k \in P} \sum_{j \in P'} X_{kj,PP'}} \right\}_{i \in P}$$

Main output

Produce vertical bars graph where $n_{PP'}$ is in the Y-axis. The X-axis would have numbers from 1 to 400. Each number would correspond to one interaction betwee the province of the seller and the one of the buyer (the unit of observation). The numbers in the X-axis would be ordered according to $\chi_{PP'}$. Number one would be the combination PP' with the lowest $\chi_{PP'}$ and number 400 would be the combination PP' with the highest $\chi_{PP'}$.

Note: Since there are many units of observations, the bars should be very thin. I would not mind if it looks like a continuous area (actually, this would be preferred!). We can try also with a line graph instead of the vertical bars. Whatever looks better.