**MACHINE LEARNING PS-1** (*Arni Parikh)*

SUMMARY STATISTICS 1

Ans 3:

| **Variable** | | **Mean** | **SD** |
| --- | --- | --- | --- |
|  |  |  |  |
| **1** | 401(k) participation | 2.616238e-01 | 4.395410e-01 |
| **2** | 401(k) eligibility | 3.713565e-01 | 4.831919e-01 |
| **3** | Net financial assets | 1.805153e+04 | 6.352250e+04 |
| **4** | Net non-401(k) financial assets | 1.041415e+04 | 5.602886e+04 |
| **5** | Total wealth | 6.381685e+04 | 1.115297e+05 |
| **6** | Income | 3.720062e+04 | 2.477429e+04 |
| **7** | Age | 4.106021e+01 | 1.034450e+01 |
| **8** | Family size | 2.865860e+00 | 1.538937e+00 |
| **9** | Married | 6.048411e-01 | 4.889094e-01 |
| **10** | IRA participation | 3.462872e+03 | 9.648028e+03 |
| **11** | Defined benefit pension | 2.710035e-01 | 4.445003e-01 |
| **12** | Home ownership | 6.351992e-01 | 4.813985e-01 |
| **13** | No high school | 1.282905e-01 | 3.344298e-01 |
| **14** | High school | 3.768028e-01 | 4.846093e-01 |
| **15** | Some college | 2.444781e-01 | 4.297990e-01 |
| **16** | College | 2.504286e-01 | 4.332817e-0 |

Ans 4:

| **Variable** | | **Mean** | **SD** | **Median** |
| --- | --- | --- | --- | --- |
|  |  |  |  |  |
| **1** | 401(k) participation | 2.616238e-01 | 4.395410e-01 | *NA* |
| **2** | 401(k) eligibility | 3.713565e-01 | 4.831919e-01 | *NA* |
| **3** | Net financial assets | 1.805153e+04 | 6.352250e+04 | 1499 |
| **4** | Net non-401(k) financial assets | 1.041415e+04 | 5.602886e+04 | 270 |
| **5** | Total wealth | 6.381685e+04 | 1.115297e+05 | 25100 |
| **6** | Income | 3.720062e+04 | 2.477429e+04 | *NA* |
| **7** | Age | 4.106021e+01 | 1.034450e+01 | *NA* |
| **8** | Family size | 2.865860e+00 | 1.538937e+00 | *NA* |
| **9** | Married | 6.048411e-01 | 4.889094e-01 | *NA* |
| **10** | IRA participation | 3.462872e+03 | 9.648028e+03 | *NA* |
| **11** | Defined benefit pension | 2.710035e-01 | 4.445003e-01 | *NA* |
| **12** | Home ownership | 6.351992e-01 | 4.813985e-01 | *NA* |
| **13** | No high school | 1.282905e-01 | 3.344298e-01 | *NA* |
| **14** | High school | 3.768028e-01 | 4.846093e-01 | *NA* |
| **15** | Some college | 2.444781e-01 | 4.297990e-01 | *NA* |
| **16** | College | 2.504286e-01 | 4.332817e-01 | *NA* |

Ans 5:

Summary Table for participants

| **Variable** | | **Mean** | **SD** | | **Median** |
| --- | --- | --- | --- | --- | --- |
|  |  |  | |  |  |
| **1** | 401(k) participation | 1.000000e+00 | | 0.000000e+00 | *NA* |
| **2** | 401(k) eligibility | 1.000000e+00 | | 0.000000e+00 | *NA* |
| **3** | Net financial assets | 3.826206e+04 | | 7.908754e+04 | 15249.0 |
| **4** | Net non-401(k) financial assets | 1.727011e+04 | | 6.587490e+04 | 2041.0 |
| **5** | Total wealth | 9.692010e+04 | | 1.277896e+05 | 53440.5 |
| **6** | Income | 4.936698e+04 | | 2.720844e+04 | *NA* |
| **7** | Age | 4.150964e+01 | | 9.658820e+00 | *NA* |
| **8** | Family size | 2.915960e+00 | | 1.470061e+00 | *NA* |
| **9** | Married | 6.904395e-01 | | 4.624016e-01 | *NA* |
| **10** | IRA participation | 5.505312e+03 | | 1.185122e+04 | *NA* |
| **11** | Defined benefit pension | 3.916731e-01 | | 4.882184e-01 | *NA* |
| **12** | Home ownership | 7.652274e-01 | | 4.239383e-01 | *NA* |
| **13** | No high school | 6.784888e-02 | | 2.515349e-01 | *NA* |
| **14** | High school | 3.457980e-01 | | 4.757194e-01 | *NA* |
| **15** | Some college | 4.306091e-01 | | 4.952570e-01 | *NA* |
| **16** | College | 1.557440e-01 | | 3.626824e-01 | *NA* |

Summary Table for non-participants

| **Variable** | | **Mean** | **SD** | | **Median** |
| --- | --- | --- | --- | --- | --- |
|  |  |  | |  |  |
| **1** | 401(k) participation | 0.000000e+00 | | 0.000000e+00 | *NA* |
| **2** | 401(k) eligibility | 1.486136e-01 | | 3.557315e-01 | *NA* |
| **3** | Net financial assets | 1.089048e+04 | | 5.525683e+04 | 200 |
| **4** | Net non-401(k) financial assets | 7.984924e+03 | | 5.188378e+04 | 57 |
| **5** | Total wealth | 5.208760e+04 | | 1.026460e+05 | 16645 |
| **6** | Income | 3.288980e+04 | | 2.231561e+04 | *NA* |
| **7** | Age | 4.090097e+01 | | 1.057285e+01 | *NA* |
| **8** | Family size | 2.848108e+00 | | 1.562325e+00 | *NA* |
| **9** | Married | 5.745117e-01 | | 4.944506e-01 | *NA* |
| **10** | IRA participation | 2.739188e+03 | | 8.620675e+03 | *NA* |
| **11** | Defined benefit pension | 2.282475e-01 | | 4.197316e-01 | *NA* |
| **12** | Home ownership | 5.891272e-01 | | 4.920258e-01 | *NA* |
| **13** | No high school | 1.497063e-01 | | 3.568077e-01 | *NA* |
| **14** | High school | 3.877886e-01 | | 4.872792e-01 | *NA* |
| **15** | Some college | 3.585576e-01 | | 4.796097e-01 | *NA* |
| **16** | College | 1.039475e-01 | | 3.052133e-01 | *NA* |

Ans 6:

Summary Table for eligible

| **Variable** | | **Mean** | **SD** | **Median** |
| --- | --- | --- | --- | --- |
|  |  |  |  |  |
| **1** | 401(k) participation | 1.000000e+00 | 0.000000e+00 | *NA* |
| **2** | 401(k) eligibility | 1.000000e+00 | 0.000000e+00 | *NA* |
| **3** | Net financial assets | 3.826206e+04 | 7.908754e+04 | 15249.0 |
| **4** | Net non-401(k) financial assets | 1.727011e+04 | 6.587490e+04 | 2041.0 |
| **5** | Total wealth | 9.692010e+04 | 1.277896e+05 | 53440.5 |
| **6** | Income | 4.936698e+04 | 2.720844e+04 | *NA* |
| **7** | Age | 4.150964e+01 | 9.658820e+00 | *NA* |
| **8** | Family size | 2.915960e+00 | 1.470061e+00 | *NA* |
| **9** | Married | 6.904395e-01 | 4.624016e-01 | *NA* |
| **10** | IRA participation | 5.505312e+03 | 1.185122e+04 | *NA* |
| **11** | Defined benefit pension | 3.916731e-01 | 4.882184e-01 | *NA* |
| **12** | Home ownership | 7.652274e-01 | 4.239383e-01 | *NA* |
| **13** | No high school | 6.784888e-02 | 2.515349e-01 | *NA* |
| **14** | High school | 3.457980e-01 | 4.757194e-01 | *NA* |
| **15** | Some college | 4.306091e-01 | 4.952570e-01 | *NA* |
| **16** | College | 1.557440e-01 | 3.626824e-01 | *NA* |

Summary Table for non-eligible

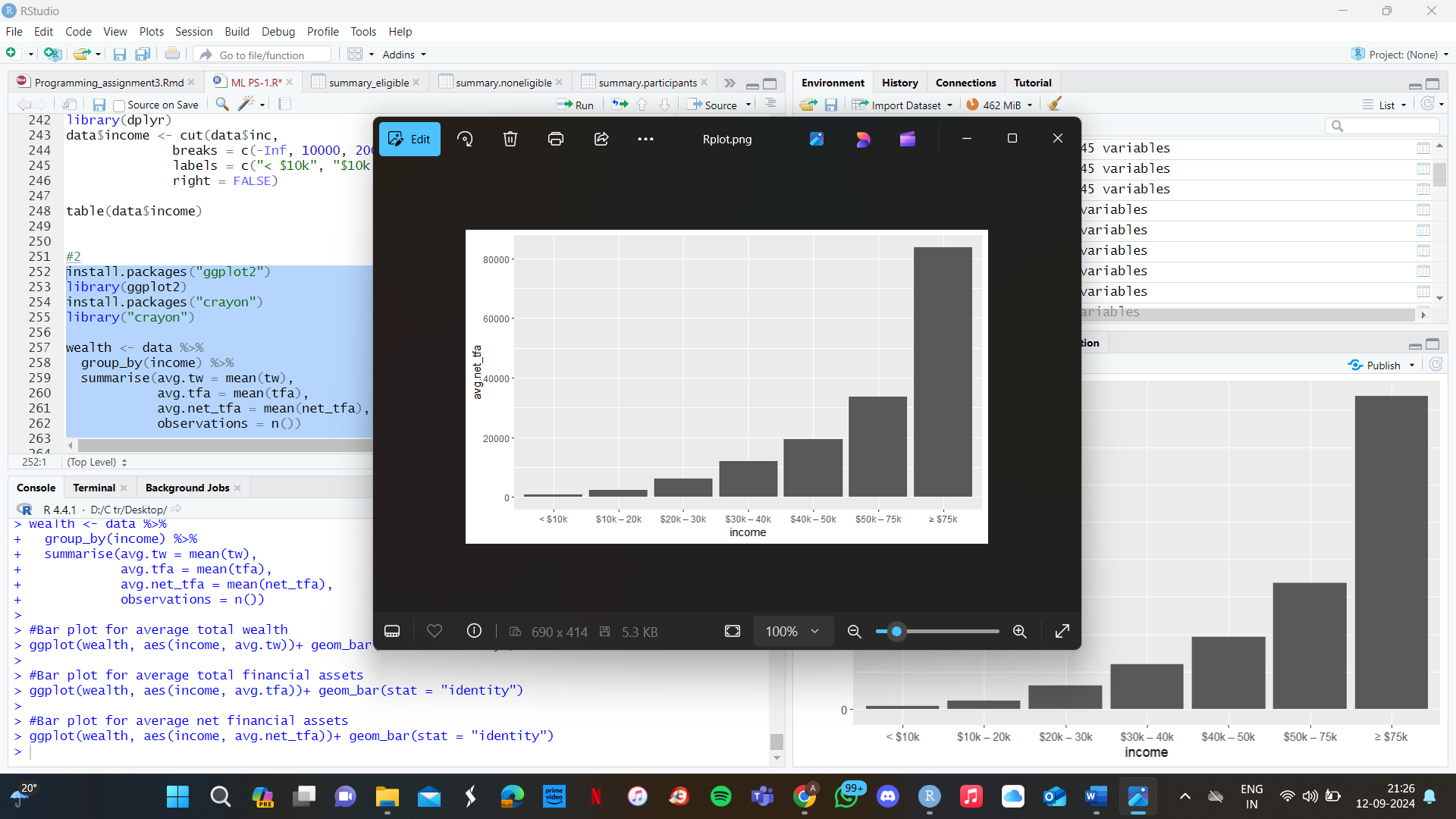
| **Variable** | | **Mean** | **SD** | **Median** |
| --- | --- | --- | --- | --- |
|  |  |  |  |  |
| **1** | 401(k) participation | 0.000000e+00 | 0.000000e+00 | *NA* |
| **2** | 401(k) eligibility | 1.486136e-01 | 3.557315e-01 | *NA* |
| **3** | Net financial assets | 1.089048e+04 | 5.525683e+04 | 200 |
| **4** | Net non-401(k) financial assets | 7.984924e+03 | 5.188378e+04 | 57 |
| **5** | Total wealth | 5.208760e+04 | 1.026460e+05 | 16645 |
| **6** | Income | 3.288980e+04 | 2.231561e+04 | *NA* |
| **7** | Age | 4.090097e+01 | 1.057285e+01 | *NA* |
| **8** | Family size | 2.848108e+00 | 1.562325e+00 | *NA* |
| **9** | Married | 5.745117e-01 | 4.944506e-01 | *NA* |
| **10** | IRA participation | 2.739188e+03 | 8.620675e+03 | *NA* |
| **11** | Defined benefit pension | 2.282475e-01 | 4.197316e-01 | *NA* |
| **12** | Home ownership | 5.891272e-01 | 4.920258e-01 | *NA* |
| **13** | No high school | 1.497063e-01 | 3.568077e-01 | *NA* |
| **14** | High school | 3.877886e-01 | 4.872792e-01 | *NA* |
| **15** | Some college | 3.585576e-01 | 4.796097e-01 | *NA* |
| **16** | College | 1.039475e-01 | 3.052133e-01 | *NA* |

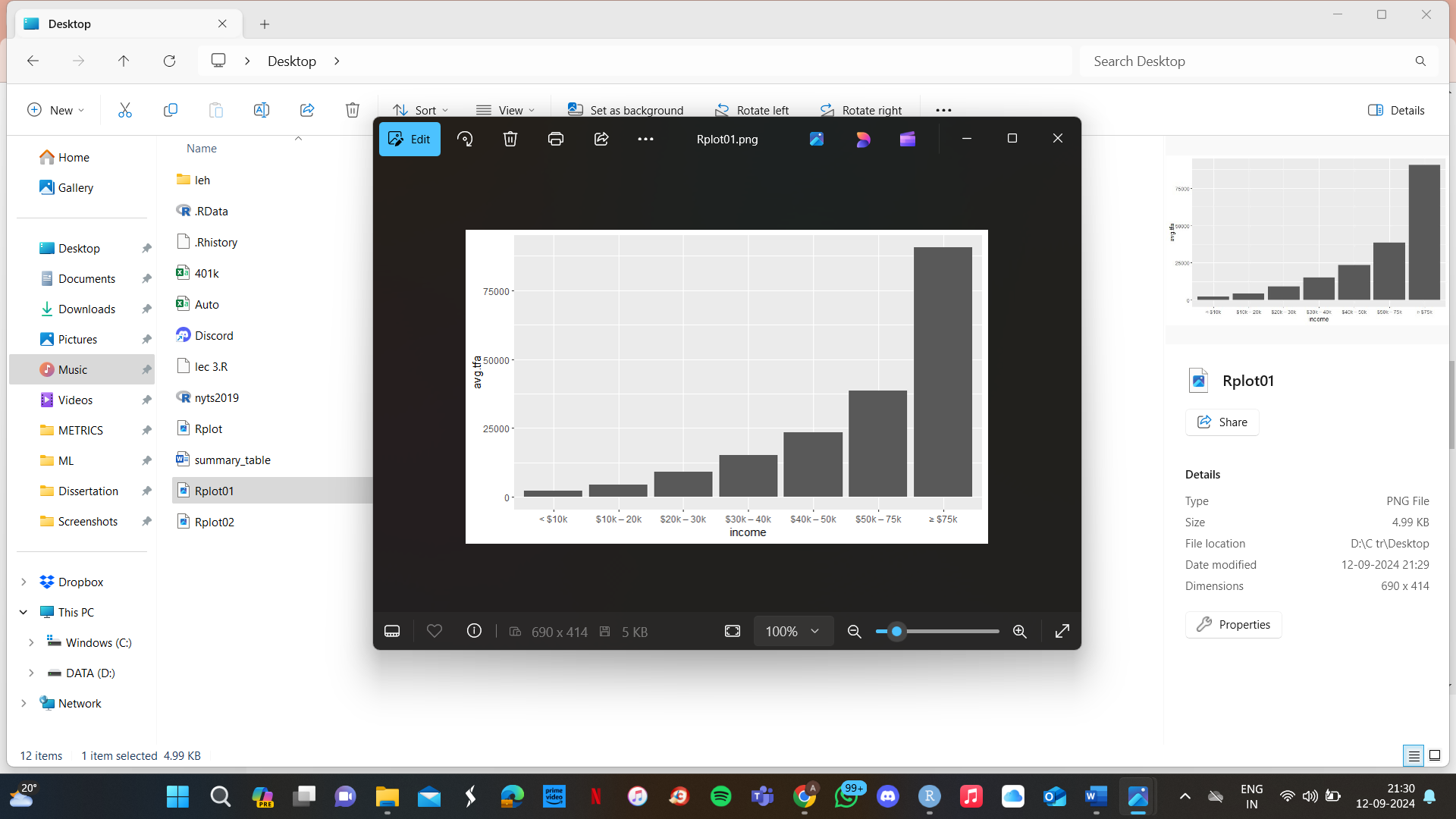
Ans 7:

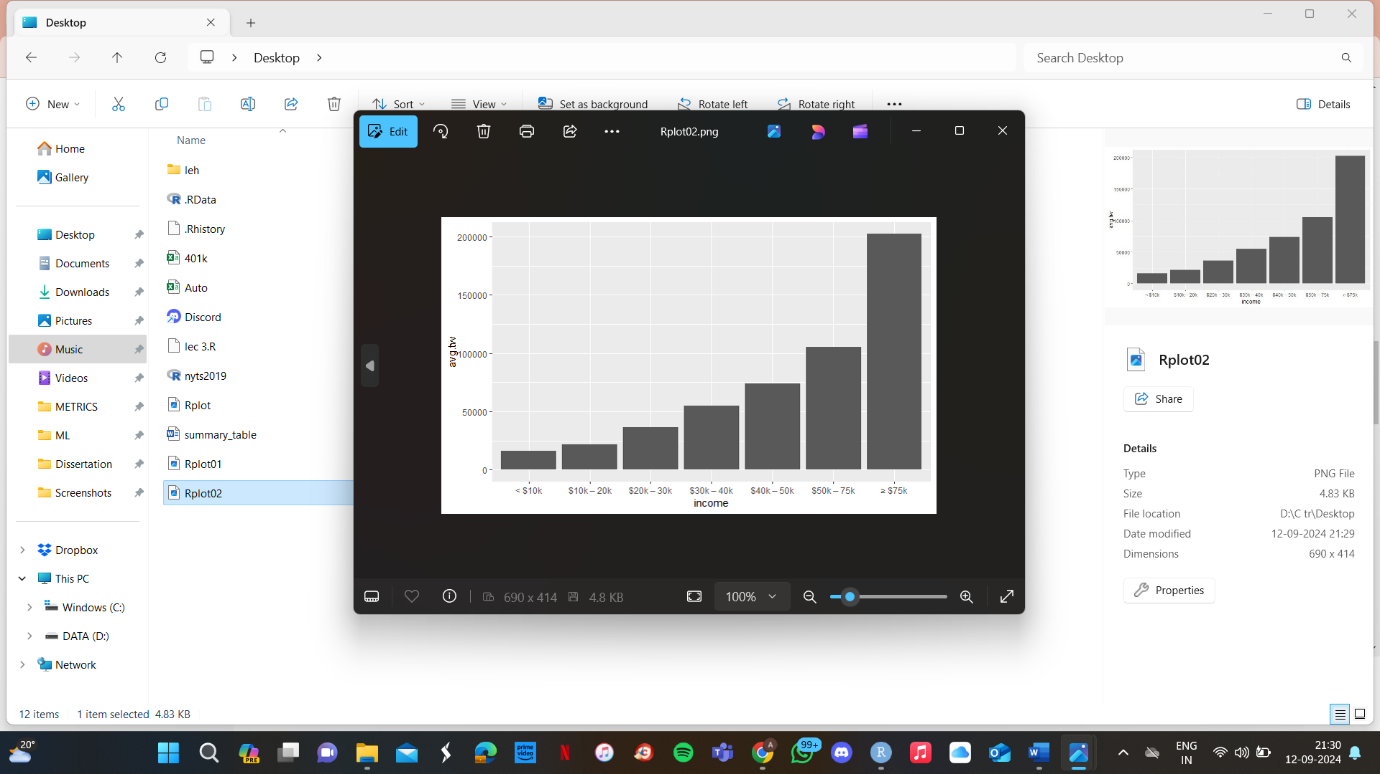
1. Median in general provides a better value for the measurement of central tendency when compared to mean for variables such as wealth since it does not account for outliers giving us a better estimate of the central tendency.
2. The association between education and 401(k) participation and eligibility indicates that higher education levels and participation and eligibility rates are directly proportionate; higher education levels show higher rates while lower education levels show lower rates. This implies education may influence financial access and decision-making.
3. The difference in net financial assets conditional on 401(k) participation ($38,262) versus non-participation ($10,890) is approximately equal to ($ 27,372) which is significantly large. These numbers might not reflect the causal effect since a lot of variables such as generational wealth, income, education etc influencing net financial assets and participation go unaccounted for leading to potential bias.

SUMMARY STATISTICS 2

Ans 2:

:





* The analysis of income versus average net financial assets reveals a clear trend: as income increases, so do net financial assets. This is evident from the steady rise in net financial assets across income brackets, with individuals earning $75k or more having notably higher net assets compared to lower income groups. This pattern suggests that higher income contributes to greater savings and investments, along with improved debt management, leading to a stronger overall financial position.
* The analysis of income versus average financial assets, which includes assets like stocks, bonds, and savings without accounting for liabilities, shows a similar trend. As income rises, so does the accumulation of financial assets. This pattern indicates that individuals with higher incomes generally have more disposable income available for investment, leading to greater financial assets in these higher income groups.
* The analysis of income versus total wealth, which encompasses both financial and non-financial assets (like property) after liabilities, demonstrates a notable increase in wealth with rising income. The sharp rise in total wealth among individuals earning $75k or more highlights that those in higher income brackets accumulate wealth through a combination of financial investments and potentially significant non-financial assets, such as real estate.

OBSERVATIONS IN EACH OF THE INCOME CATEGORIES:

| **income** | | **count** | **avg\_tfa** | |
| --- | --- | --- | --- | --- |
|  |  |  |  |
| **1** | < $10k | 638 | 2380.895 |
| **2** | $10k – 20k | 1948 | 4485.763 |
| **3** | $20k – 30k | 2074 | 9188.871 |
| **4** | $30k – 40k | 1712 | 15210.471 |
| **5** | $40k – 50k | 1204 | 23477.452 |
| **6** | $50k – 75k | 1572 | 38658.823 |
| **7** | ≥ $75k | 767 | 90525.845 |

| **income** | | **count** | **avg\_net\_tfa** |
| --- | --- | --- | --- |
|  |  |  |  |
| **1** | < $10k | 638 | 734.6959 |
| **2** | $10k – 20k | 1948 | 2308.4584 |
| **3** | $20k – 30k | 2074 | 6311.1249 |
| **4** | $30k – 40k | 1712 | 11937.6636 |
| **5** | $40k – 50k | 1204 | 19347.9502 |
| **6** | $50k – 75k | 1572 | 33707.9198 |
| **7** | ≥ $75k | 767 | 83709.2829 |

| **income** | | | **count** | | **avg\_tw** |
| --- | --- | --- | --- | --- | --- |
|  |  |  | |  | |
| **1** | < $10k | 638 | | 16235.33 | |
|  | $10k – 20k | 1948 | | 21620.00 | |
| **3** | $20k – 30k | 2074 | | 36730.19 | |
| **4** | $30k – 40k | 1712 | | 55118.71 | |
| **5** | $40k – 50k | 1204 | | 74005.50 | |
| **6** | $50k – 75k | 1572 | | 105284.90 | |
| **7** | ≥ $75k | 767 | | 202239.87 | |

Ans 3:

|  | **e401** | **Participation Rate** |
| --- | --- | --- |
| **1** | 0 | 0.00000 |
| **2** | 1 | 70.45084 |

Shows how many eligible individuals participate in 401(k) compared to ineligible individuals. (Participation by Eligibility)

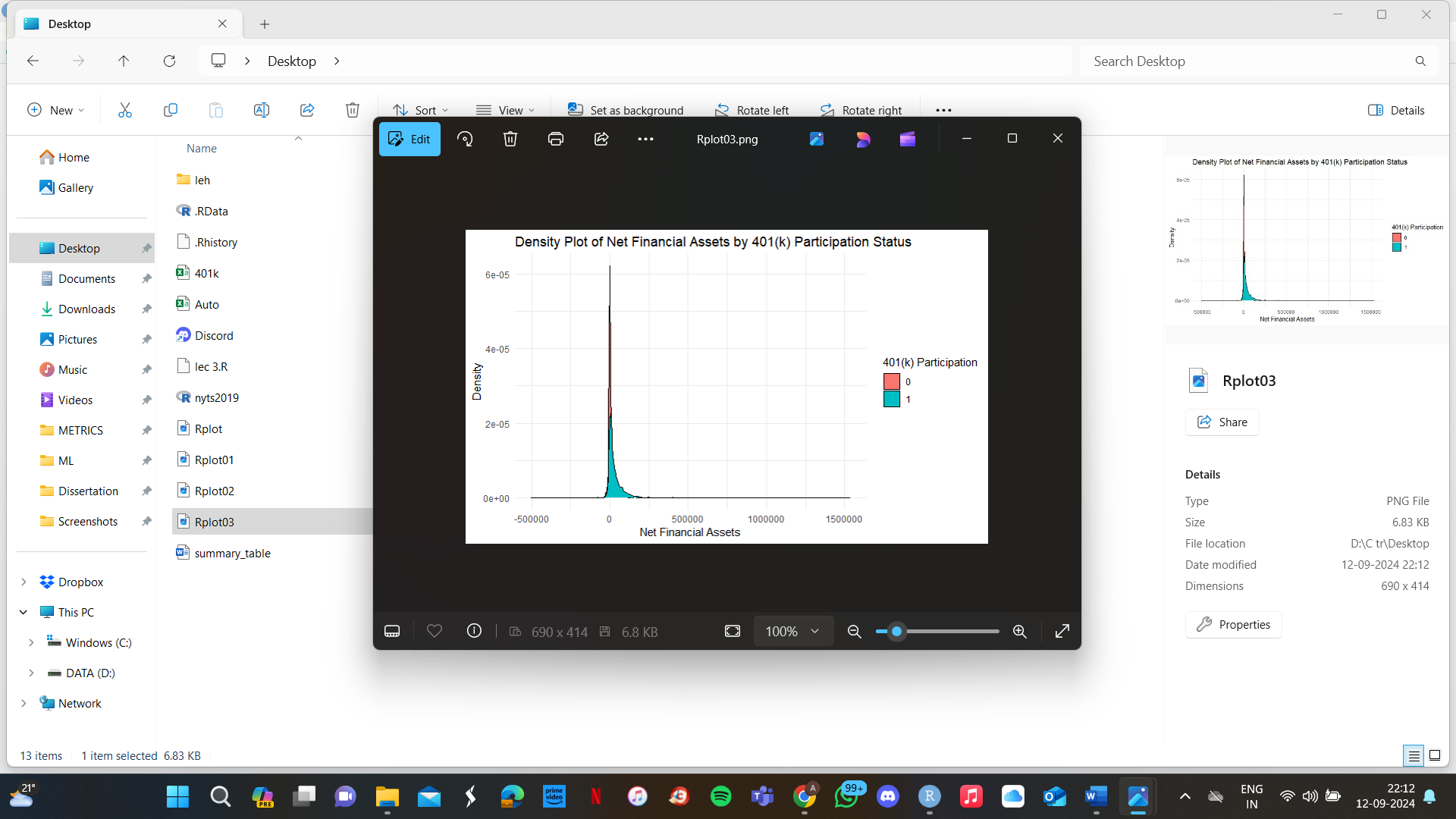
|  | **e401** | **Participation Rate** |
| --- | --- | --- |
| **1** | 0 | 14.86136 |
| **2** | 1 | 100.00000 |

Shows how many participants are eligible for 401(k) compared to non-participants. (Eligibility by Participation)

Ans 4:

The ratios reveal that no one ineligible for a 401(k) participates in the program, while a significant 70.45% of eligible individuals do. Conversely, all participants are eligible for a 401(k), but only 14.86% of non-participants are eligible. These findings align with expectations: participation is high among those eligible, and eligibility is exclusive to participants, highlighting that 401(k) participation is strongly linked to eligibility and that most non-participants are not eligible.

Ans 5:



The density plot shows that most people, whether they participate in a 401(k) or not, have low financial assets, with many having close to nothing. Both groups have a few individuals with much higher financial assets, but most are on the lower end. People who participate in a 401(k) tend to have more financial assets than those who do not, but the difference isn’t huge, especially for those with smaller amounts of assets. Both groups exhibit a right-skewed distribution, indicating that a small subset of individuals has significantly higher financial assets.