Answers 6.2 Task

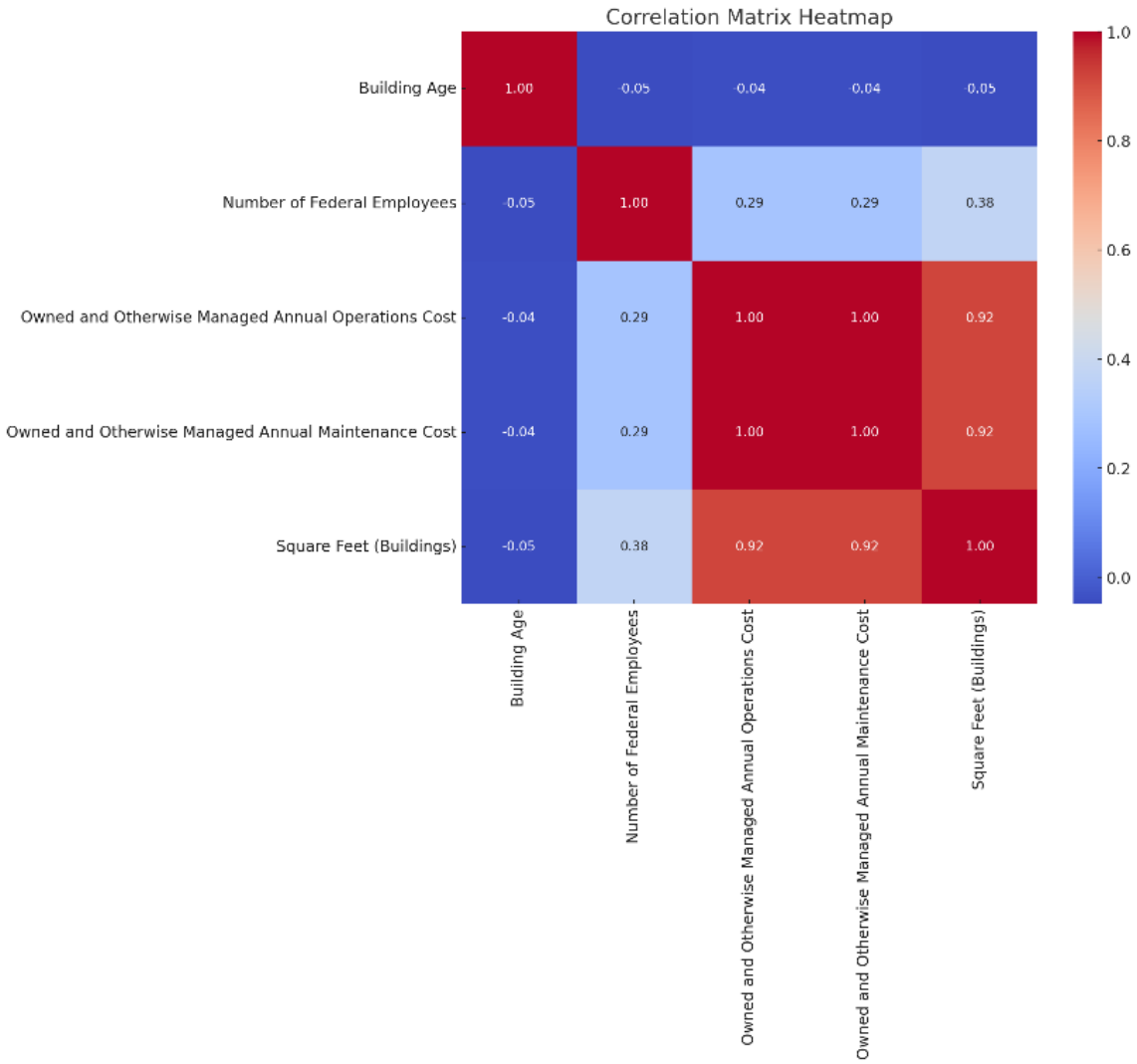
Q2. Use the questions you defined in the previous task to pick out variables from your data set suitable for your exploratory visual analysis.

My focus for this task is on space utilization: **How to measure the efficiency of office space?** In analyzing office space utilization for a government setting, I selected variables that provide insights into space efficiency and utilization.

The chosen variables include **Real Property Type and Real Property Use** to categorize the types and purposes of properties, respectively. **Utilization status** helps identify how well the space is being used, while **Building Age** offers context on the property's lifecycle. The **Number of Federal Employees** provides a measure of density, which, along with **Square Feet (Buildings)**, can be used to assess space allocation efficiency. Additionally, **Annual Operations Cost and Annual Maintenance Cost** are crucial for evaluating financial efficiency in maintaining these spaces. These variables collectively allow for a comprehensive analysis of both physical and economic aspects of space utilization.

Q3. Create a correlation matrix heatmap (colored). Discuss what the coefficients in the plot mean in terms of the relationships between the variables.

The correlation matrix heatmap displays the pairwise correlation coefficients between the selected numeric variables. The coefficients range from -1 to 1, where values close to 1 indicate a strong positive correlation, values close to -1 indicate a strong negative correlation, and values around 0 suggest little to no linear relationship.



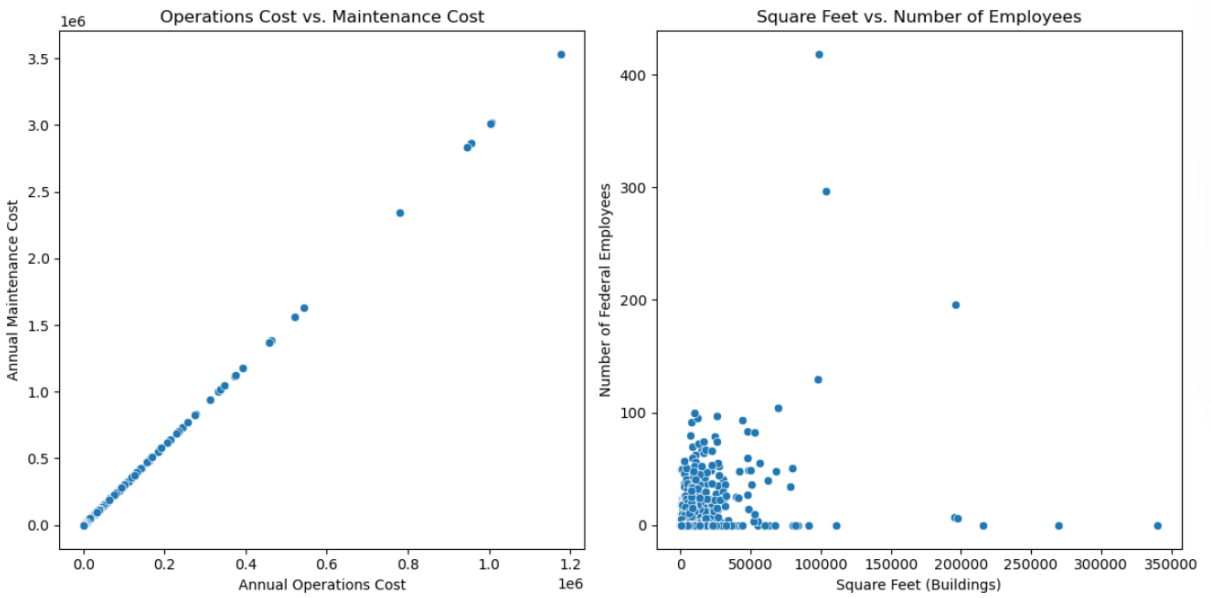
In the heatmap reveals:

- **Building Age and Number of Federal Employees** show a near-zero correlation, indicating no significant relationship between the age of the building and the number of employees.
- **Square Feet (Buildings) and Number of Federal Employees** have a positive correlation, suggesting that larger buildings generally accommodate more employees.
- **Owned and Otherwise Managed Annual Operations Cost and Maintenance Cost** are strongly positively correlated, indicating that buildings with higher operational costs tend also to have higher maintenance costs, likely due to their size or complexity.
- **Square Feet (Buildings) and Annual Costs (both Operations and Maintenance)** also show positive correlations, which is expected since larger spaces typically incur higher costs.

Q4. Create a scatterplot (or plots) for the variables with the strongest correlations and examine the nature of their relationships. Discuss the outputs.

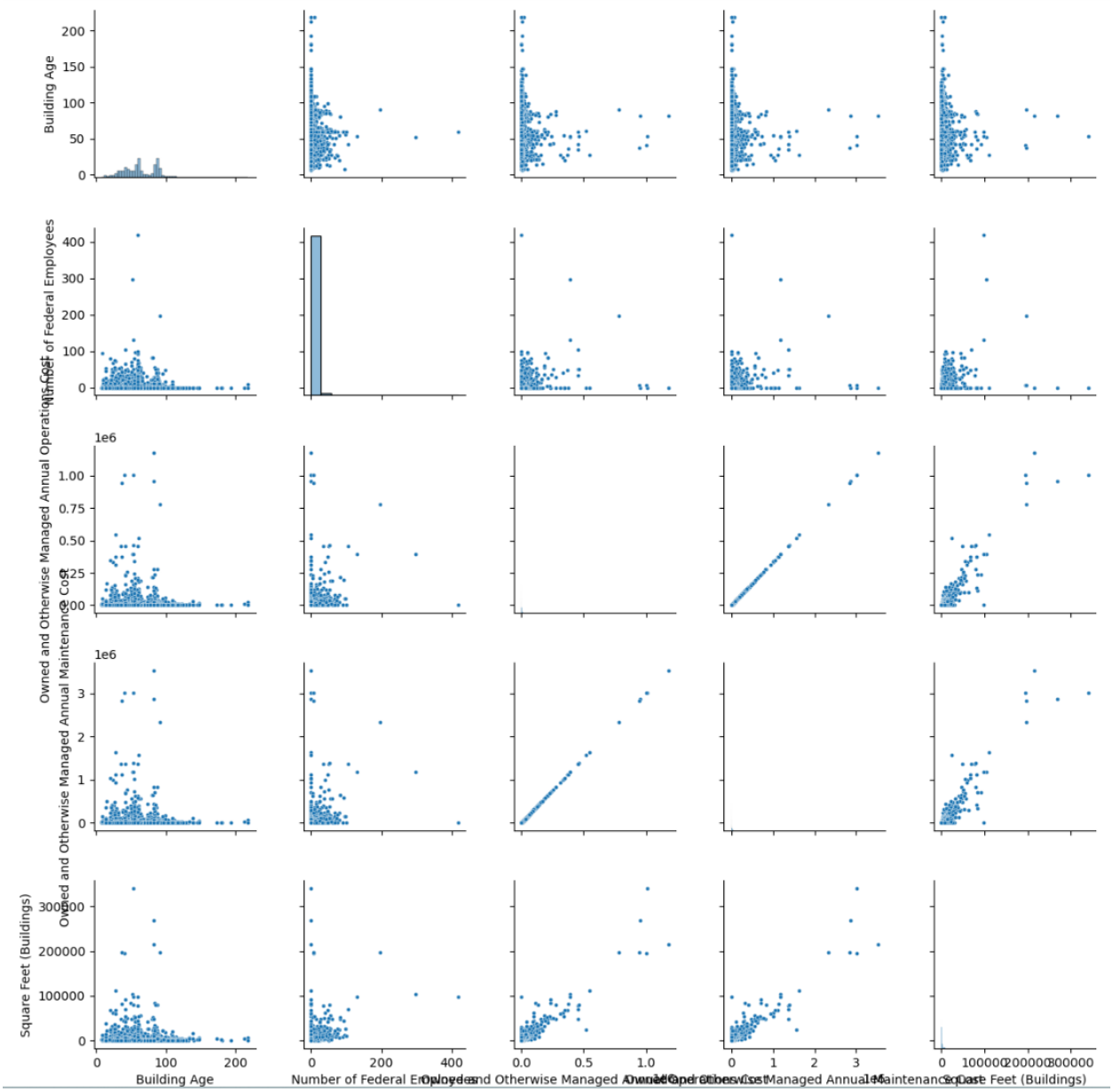
The scatterplots illustrate the relationships between key variables in the dataset. The first plot, showing the relationship between annual operations and maintenance costs, reveals a strong positive correlation, indicating that buildings with higher operational expenses also tend to incur greater maintenance costs.

The second plot depicts square footage versus the number of federal employees and suggests a general trend where larger buildings house more employees. However, the variation in data points indicates that the relationship is not strictly linear, suggesting possible underutilization in larger spaces or intensive use in smaller ones. These insights are valuable for understanding how building size and associated costs relate to the number of employees and overall space utilization in government offices.



Q5. Create a pair plot of the entire data set. Comment on the distribution of the variables and mark variables you'd like to explore further with an explanation of why.

Since the pair plot provides a visual overview of the distributions and relationships between the numeric variables in the dataset, the diagonal plots show the distribution of each variable, revealing that some variables, such as the Annual Operations Cost and Annual Maintenance Cost, may have skewed distributions, possibly indicating a few buildings with significantly higher costs than others. The off-diagonal plots display scatterplots for each pair of variables, highlighting their relationships.

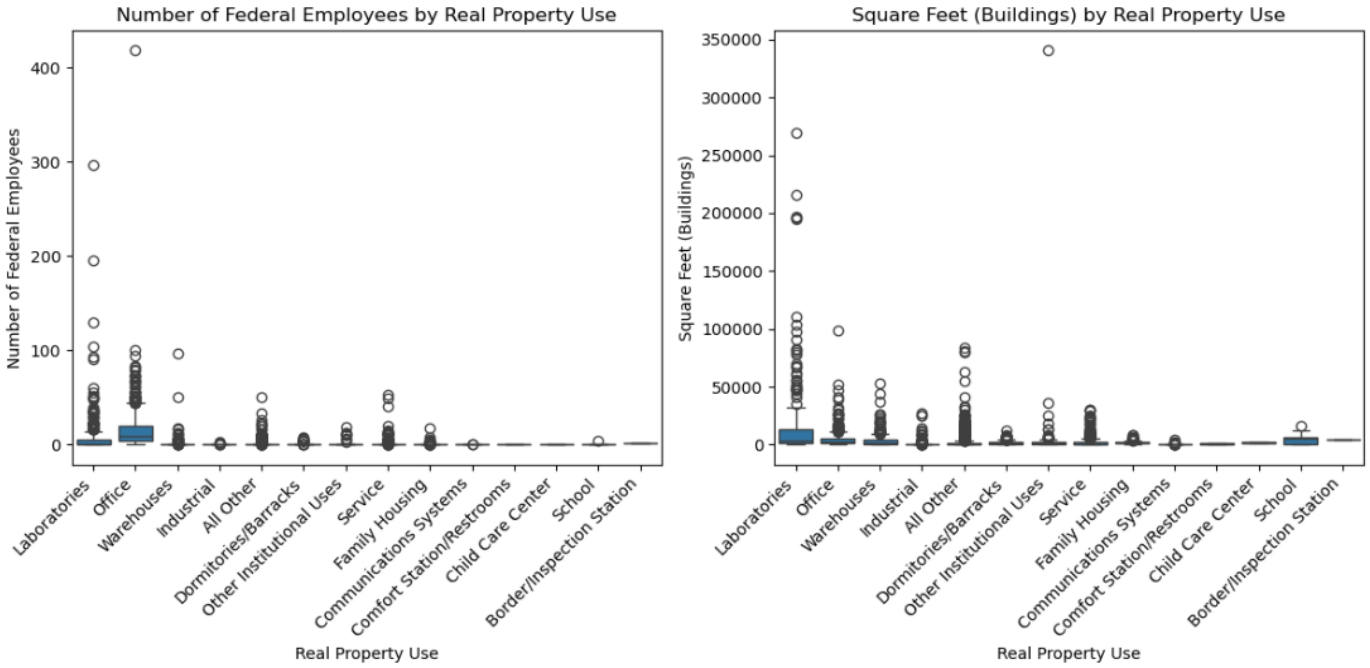


Some observations and variables to explore further:

- **Square Feet (Buildings) vs. Number of Federal Employees:** Positive correlation with significant variance; further exploration can reveal insights into space utilization and employee density.
- **Annual Operations Cost and Annual Maintenance Cost:** Strong linear relationship; investigating this can help understand cost drivers and identify potential areas for optimization.
- **Building Age:** Uniform distribution; exploring its impact on maintenance needs and operational efficiency can provide valuable insights into lifecycle costs and utilization.

Q6. Create a categorical plot and interpret the results.

To create a categorical plot, used Real Property Use as the categorical variable and analyze how it relates to Number of Federal Employees and Square Feet (Buildings). The box plots below provided a visual representation of the distribution of the number of federal employees and building size (in square feet) across different property uses.



- **Number of Federal Employees by Real Property Use:** The box plot shows the distribution of employees for each property use type. "Office" properties have a wider range of employees, indicating variability in office sizes. Other property uses, such as "Laboratories" and "Storage," tend to have a more consistent number of employees, as indicated by the smaller interquartile ranges.

- **Square Feet (Buildings) by Real Property Use:** This plot shows the distribution of building sizes for different property uses. "Office" and "Storage" properties generally occupy larger spaces, as indicated by the higher median and range of square footage. On the other hand, "Laboratories" tend to have a smaller and more consistent building size.

Q7. Revisit the questions you generated in the previous task and write answers to those you can based on the exploration you've conducted so far. Add any new questions that may have arisen based on the early findings in your visual exploration.
Based on the previously generated questions and exploratory visual analysis, below are my findings:

- **Space Optimization:** There are varied levels of space utilization across property types, particularly in office spaces, indicating potential underutilization. Further analysis could identify areas for repurposing or consolidation to improve efficiency.

- **Utilization Efficiency:** Different property types show varying levels of utilization, with office buildings having the most variability. This suggests that some properties may not be used as efficiently as others.

- **Occupancy Trends:** There is a positive correlation between building size and the number of employees, but with significant variability, indicating that some large spaces may be underutilized.

New Questions for Further Exploration

- **Cost Efficiency:** How do operational and maintenance costs correlate with building size and utilization? Are there opportunities for cost savings in properties with high costs but low utilization?

- **Regional Analysis**: Are there regional differences in the utilization and cost efficiency of federal properties? Identifying regional patterns can help target specific areas for improvement.

Q8. Define any hypotheses that you can at this point. You'll be testing these later on.

Based on my analysis conducted so far, below are potential hypotheses for further testing:

Hypothesis 1: Underutilized Office Spaces
Hypothesis: Office buildings with larger square footage have lower utilization rates (measured by the number of federal employees per square foot) compared to smaller office spaces.
Rationale: Preliminary observations suggest variability in utilization rates, especially in larger office buildings.

Hypothesis 2: Cost Correlation with Utilization
Hypothesis: Properties with higher operational and maintenance costs have lower utilization rates, indicating potential inefficiencies.
Rationale: Observations show a strong relationship between costs and building size, but the correlation with utilization efficiency needs to be tested.

Hypothesis 3: Regional Differences in Utilization
Hypothesis: There are significant regional differences in the utilization efficiency of federal properties, with some regions consistently showing higher or lower utilization rates.
Rationale: Regional patterns in property management practices may affect utilization rates.