Name: Minh Khoi Ngo, Student ID: 30889472, Teaching Associate: Michael Niemann, Applied 12

PE2: Water Resource Management and Sustainability in Arid Regions

Introduction

Topic Detail

In the arid and semi-arid regions of the world, sustainable water resource management is not merely a policy option but a vital necessity (Liu, Tang, Zhao, & Wang, 2024). Given the negative impacts of climate change and the inherent water shortages in these arid regions (Liu, Tang, Zhao, & Wang, 2024), it has created challenges as well as opportunities for innovation in water management. This project focuses on evaluating the effectiveness of various technologies and policies aimed at enhancing water sustainability in these critical areas.

Problem Description

Arid regions are characterized by their limited water resources, which are increasingly under pressure from factors such as population growth, agricultural demands, and climatic variability (Zhang et al., 2024). Effective management strategies are crucial to ensure the equitable and sustainable use of these scarce resources (Magdy, 2024). Understanding the impact of different management technologies and policies can help identify best practices and areas needing improvement.

Research Questions

This report addresses three fundamental questions concerning water sustainability in arid landscapes:

1. What technologies are currently being used in arid regions to enhance water sustainability, and how effective are they?
2. How have government policies in these areas impacted water management practices?
3. What are the relationships between water resource management strategies and their impacts on agricultural productivity and urban water supply?

The data source:

_ Fiscal Year Annual Data (California Department of Water Resources, 2023) used to address question 1.

_ NYC Building Energy and Water Data Disclosure for Local Law 84 (2022–Present) (NYC Mayor's Office of Sustainability, 2022) used to address question 2.

_ Water Related Land Use (2023) (Utah Division of Water Resources, 2024) used to address question 3.

Brief Motivation

My motivation for this study stems from personal observations and experiences during travels to various arid regions. Witnessing firsthand the challenges and innovations in water management has inspired me to contribute academically to this field. By identifying effective strategies and technologies, this research aims to support global efforts in achieving sustainable water management in environments most vulnerable to water scarcity.

Data Wrangling and Checking

Data Source: Fiscal Year Annual Water Dataset (2023) (California Department of Water Resources, 2023)

The dataset obtained from the California State Water Resources Control Board contains 404 rows and 7 columns.

Data Source URL: https://datasets.ai/datasets/data-for-calculating-efficient-outdoor-water-uses-147dd

Overview: This dataset contains various performance indicators for urban water suppliers across California, including Evapotranspiration (ETo), Precipitation Efficiency, and Supplier Name, allowing for comparative performance analysis (California Department of Water Resources, 2023).

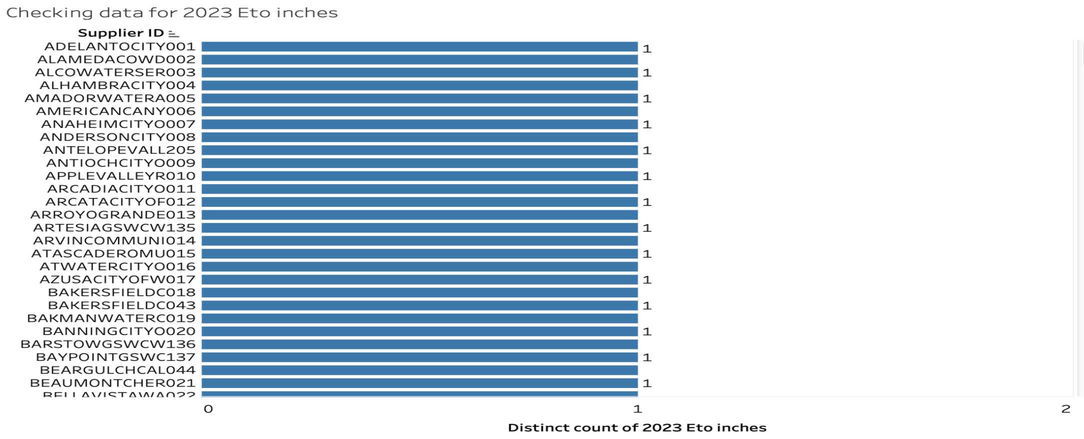

Figure 1. Supplier Id and distinct count of 2023 Eto inches

_ To verify the consistency of the 2023 ETo inches (Evapotranspiration) values, I used Tableau to plot a bar chart where the distinct count of ETo values per supplier is shown on the x-axis and each Supplier ID is listed on the y-axis. Based on the Figure 1, each supplier had exactly one unique ETo value, meaning there were no duplicates or conflicting values for this attribute. This confirms that the ETo data is complete and formatted correctly for each supplier.

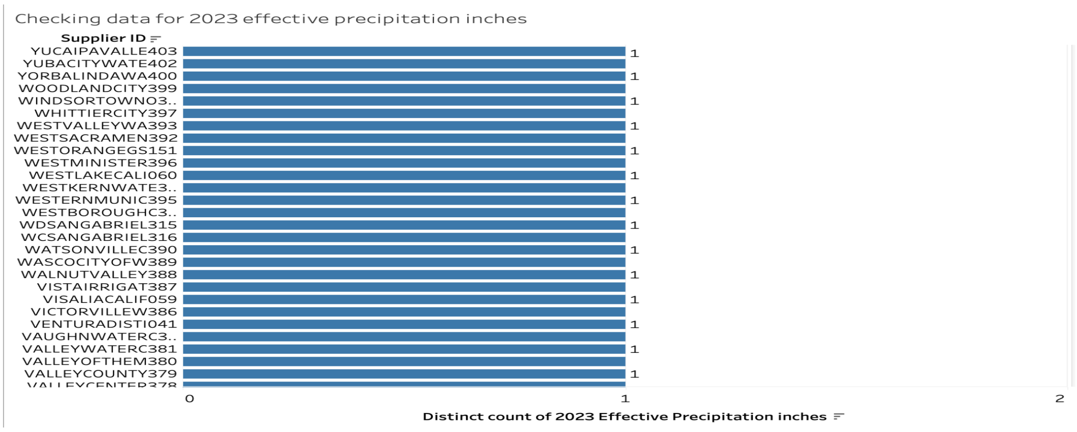

Figure 2. Supplier ID and distinct count of 2023 effective precipitation inches

_ I used the same method for checking data of 2023 effective precipitation inches. I created a bar chart by using Tableau, the visualisation displays the distinct count of 2023 effective precipitation inches on x-axis and supplier ID on y-axis. As illustrated in Figure 2, each supplier has exactly one unique value, confirming that there are no format inconsistencies in this field. This indicates that the precipitation data was consistently recorded for all water suppliers.

_ This method was applied to other key attributes. They all satisfied the data checking process.

Data Source: NYC Building Energy and Water Data Disclosure for Local Law 84 (2022–Present) (NYC Mayor's Office of Sustainability, 2022)

Overview: This dataset provides detailed building-level data on energy and water use for covered buildings in New York City with approximately 64169 rows and 265 columns (NYC Mayor's Office of Sustainability, 2022).

It includes attributes such as ENERGY STAR Score, Site EUI, Primary Property Type, and Borough (NYC Mayor's Office of Sustainability, 2022). It enables analysis of energy efficiency performance and benchmarking across building types and locations.

Data Source URL:
https://data.cityofnewyork.us/Environment/NYC-Building-Energy-and-Water-Data-Disclosure-for-/5zyy-y8am/data_preview

_ To ensure the integrity of the dataset, I used Tableau to generate a distinct count bar chart of City values for each Property ID. This helped identify if a single property had been assigned multiple city names.

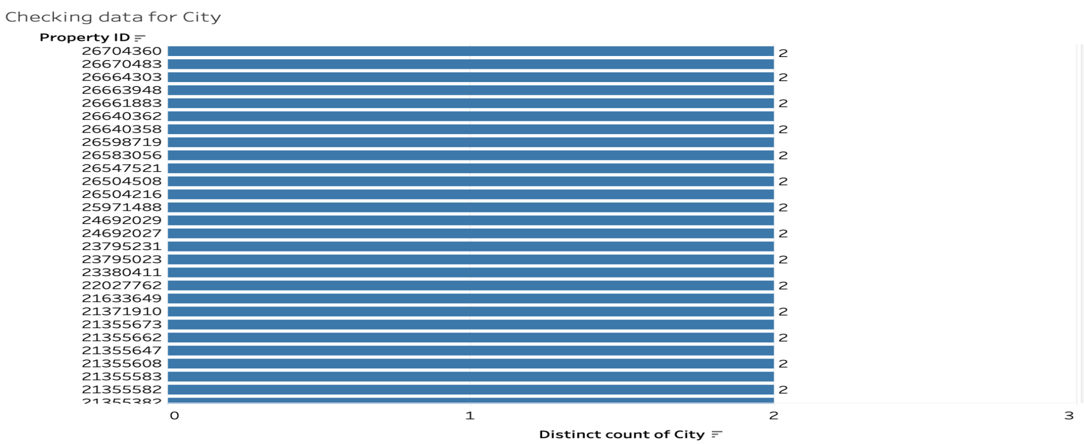

Figure 3. Property ID and distinct count of city

Based on Figure 3, several property IDs had a distinct city count of 2, indicating that the same property was listed under two different city labels. This is likely due to inconsistent naming conventions.

	Property ID	Address 1	City	Postal Code
572	6271350	66 AINSLIE ST	BROOKLYN	11211
26165	2271551	395 Dekalb Ave	Bklyn	11205
5	14377690	1680 Ocean Ave	Brooklyn	11230
18411	22895668	79 WILLOW ST	Brooklyn Heights	11201
466	11715416	1430 43rd Street	Brooklyn NY	11219
30207	8460007	844 50th Street	Brooklyn NY 11220	11220
7650	2944604	1824 53rd St.	Brooklyn,	11204
26028	25429361	3178 Nostrand Avenue	Brooklyn, NY	11229
669	8413260	124 Jackson Street	Brooklyn, New York	11211
18641	16090115	214 THOMAS S BOYLAND S	Brooklyns	11233

Figure 4. City's name with different format

_ As illustrated in Figure 4. I used python to take out all "Brooklyn" cities with different format names. These inconsistent values can negatively impact group-based analysis (e.g., average usage by borough), as each variation would be treated as a separate category.

_ I resolved this issue by normalizing all city name variations referring to Brooklyn into a single standard label, "Brooklyn". This included correcting inconsistent formats such as "BROOKLYN", "Brooklyn NY", "Bklyn", and "Brooklyn, New York" to ensure consistent representation across the dataset.

_ I applied the same method to other cities that exhibited similar inconsistencies such as New York, Queens, Flushing, and Astoria. I also normalized all variants into a consistent format. This process ensured that each city was uniformly represented across the dataset. After the cleaning, the final list of standardized unique cities included: ['New York', 'Brooklyn', 'Bronx', 'Queens', 'Staten Island', 'Jamaica', 'Flushing', 'Astoria'].

_ Another issue identified during data checking was the use of the city name "New York". While technically valid, this label is ambiguous in the context of this dataset, as it does not specify which borough the property belongs to (e.g., Manhattan, Bronx, Queens). To ensure the accuracy of borough-level analysis and avoid misclassification, I decided to exclude all rows where the city was labeled as “New York”. This helped maintain consistency and clarity in geographic comparisons across the dataset.

Data Source: Water Related Land Use (2023) (Utah Division of Water Resources, 2024)

Overview: The dataset was obtained in CSV format and contains 331,643 rows and 18 columns and provides detailed information on agricultural land use across Utah, USA (Utah Division of Water Resources, 2024). Key variables include: IRR_Method (the irrigation method used (e.g., Drip, Flood, Sprinkler)), Acres (the area of land managed under each method), Basin, County, CropGroup, and SubArea (spatial and categorical identifiers) (Utah Division of Water Resources, 2024).

Data Source URL:
https://dwre-utahdnr.opendata.arcgis.com/datasets/water-related-land-use-2023/explore?location=39.581924%2C-111.604167%2C6.70

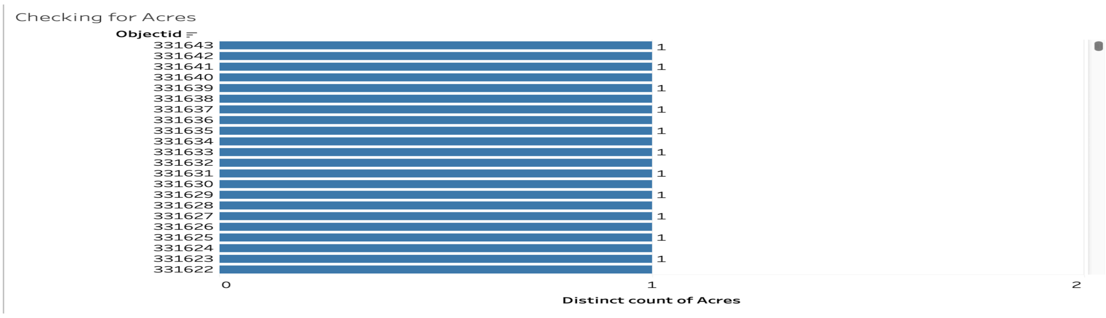

Figure 5. Distinct count of Acres by ObjectID

_ To check the outliers of Acres or the incorrect format of Acres, I used Tableau to construct a bar chart of “object id” for rows and distinct count of “Acres” as columns. Based on the figure 5, it confirms that the `Acres` field contains valid numeric entries with no formatting inconsistencies or data entry errors.
_ Using the same approach, I checked the Basin attribute to identify any irregular values.

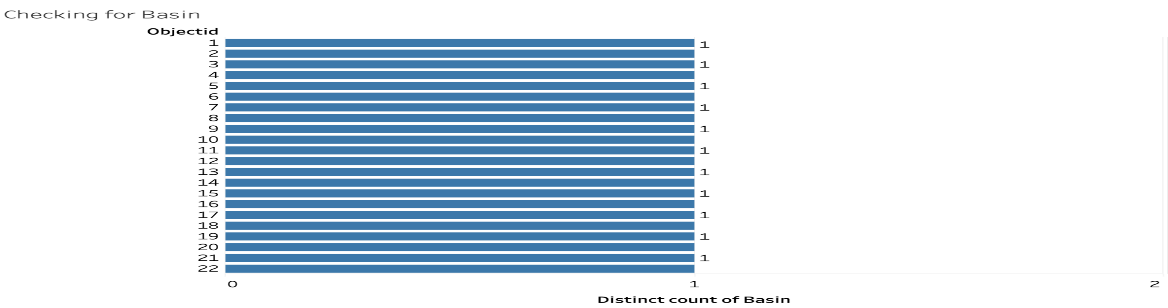

Figure 6. Distinct count of `Basin` by `ObjectID`

_ As seen in Figure 6, the count of distinct Basin values per ObjectID is consistent and returns a value of 1 in every case. This suggests the data in the Basin column is well-structured and does not contain categorical errors.

_ This method was applied to other key attributes such as IRR_Method, Land Use, and CropGroup, they passed the consistency check.These visual methods in Tableau offered a simple but effective way to validate field structure and detect potential anomalies across hundreds of thousands of records.

Data Exploration

Data Source: Fiscal Year Annual Data (California Department of Water Resources, 2023)

A horizontal bar chart was created to display the total evapotranspiration (ETo) values in inches for each supplier in 2023. Bar chart is useful to make comparisons between numerical and categorical data (Zeng & Rouse, 2022). ETo is a crucial indicator of water demand, representing the combined loss of water through evaporation and plant transpiration. High ETo values typically indicate greater need for irrigation, especially in arid regions.

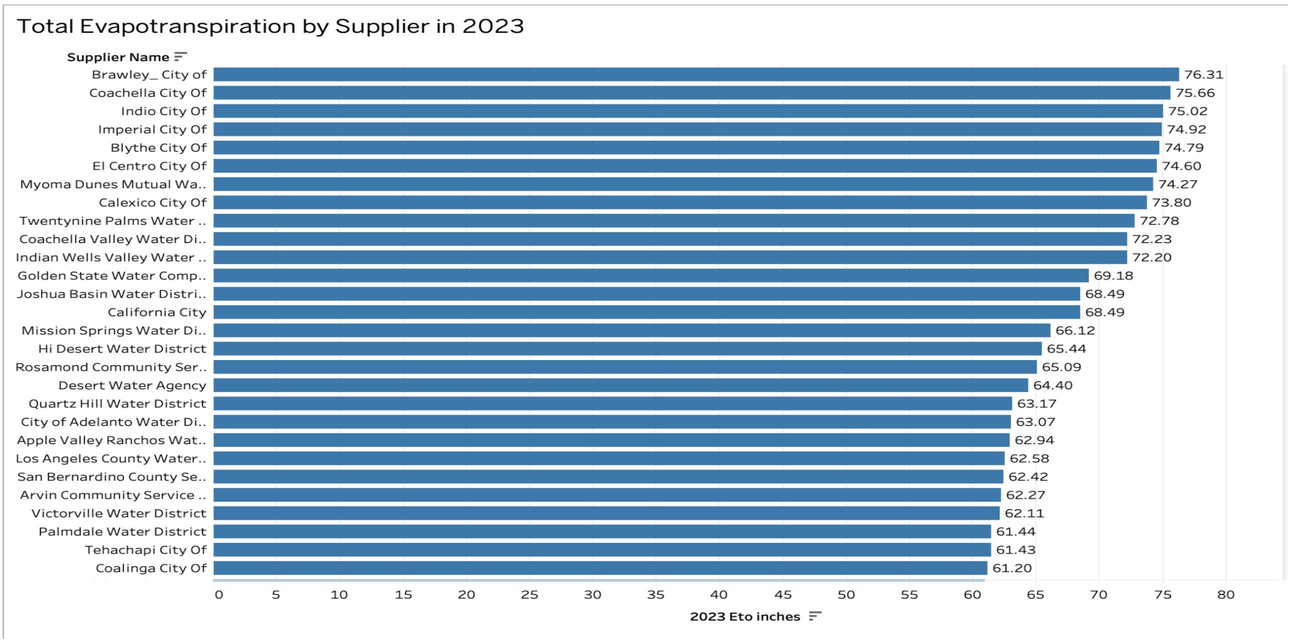

Figure 7. Total Evapotranspiration by Supplier in 2023

_ Figure 7 reveals that Brawley City, Coachella City, and Indio City are among the highest in terms of ETo, each exceeding 75 inches. These high demand regions likely require more sophisticated water management strategies to ensure sustainability. In contrast, suppliers like Los Angeles County Water show lower ETo values, indicating reduced irrigation pressure. This ranking helps identify priority areas for implementing or evaluating the effectiveness of water-saving technologies.

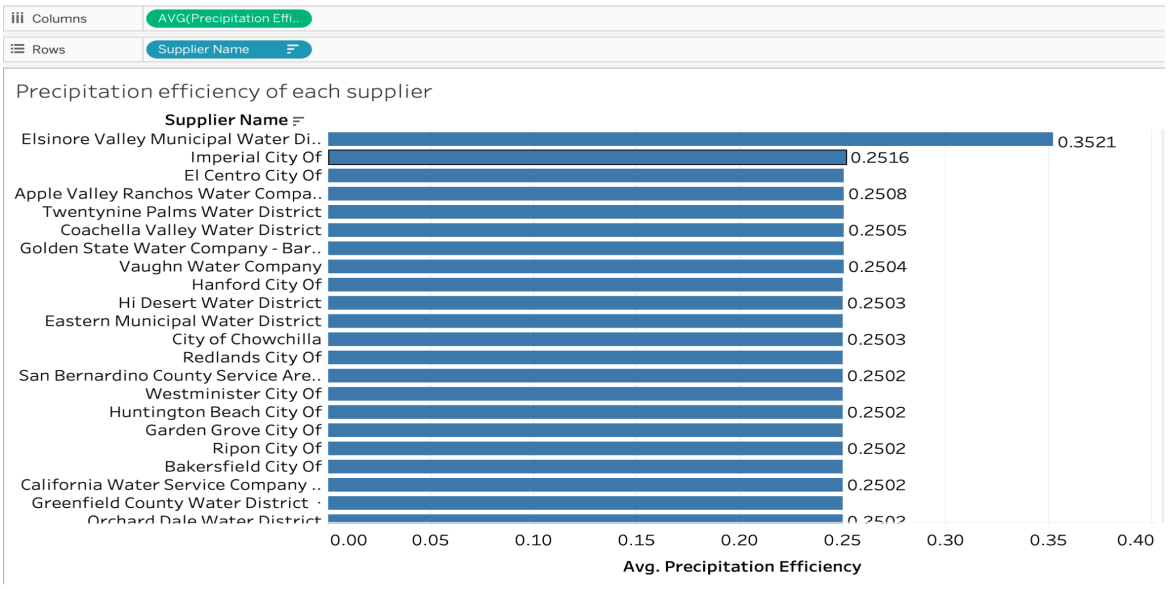

Figure 8. Average Precipitation Efficiency by Supplier

To indirectly assess how effectively different suppliers manage water in arid conditions, a new metric called Precipitation Efficiency was calculated for each supplier by creating a calculated field in tableau. This metric is defined as: Precipitation Efficiency= [2023 Effective Precipitation inches] / [2023 Total Precipitation inches] (Sustainability Directory, 2025). This ratio indicates how much of the total rainfall is actually retained and available for plant and crop use (Sustainability Directory, 2025). Higher values suggest more effective irrigation systems, soil moisture retention strategies, or urban planning that promotes water capture, all of which can be linked to the use of sustainable technologies (Sustainability Directory, 2025). The bar chart from figure 8 was created in Tableau, using AVG(Precipitation Efficiency) on x-axis to ensure that results were not inflated by repeated entries and supplier name on y-axis. Most suppliers clustered tightly around a value of 0.25. However, Elsinore Valley Municipal Water District stood out with a much higher efficiency of 0.3521, indicating significantly better rainfall utilization. This finding suggests that certain suppliers may be employing more effective technologies such as drip irrigation, smart controllers, or soil conditioning which allow them to capture and retain more water. Although the dataset does not explicitly list the technologies used, this ratio serves as a useful proxy indicator of sustainability performance.

Data Source: NYC Building Energy and Water Data Disclosure for Local Law 84 data (NYC Mayor's Office of Sustainability, 2022)

To evaluate whether government policies such as Local Law 84 are influencing energy efficiency practices across different areas of New York City, a bar chart was created by using Tableau showing the average of ENERGY STAR Score (on y-axis) by city (on x-axis). The ENERGY STAR Score (ranging from 0 to 100) serves as a standardized performance metric for energy efficiency in buildings, where higher scores indicate better efficiency. Before visualizing the data, I converted the data type of the ENERGY STAR Score from string to number (whole) in Tableau to ensure accurate aggregation and analysis.

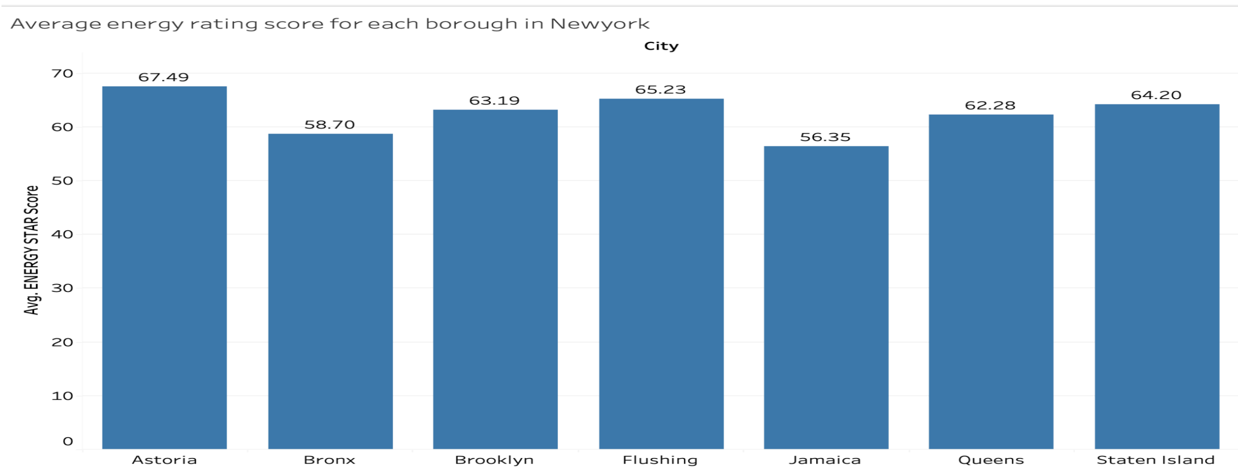

Figure 9. Average ENERGY STAR Score by different areas of New York city

_ From Figure 9, Astoria (67.49) and Flushing (65.23) had the highest average scores, suggesting effective local management practices or newer, more efficient infrastructure. In addition, Jamaica (56.35) and Bronx (58.70) had the lowest scores, which could indicate aging infrastructure, reduced investment in upgrades, or challenges in enforcement of Local Law 84. Brooklyn, Queens, and Staten Island displayed mid-range performance (62–64), highlighting areas with balanced efficiency strategies.

_ These patterns suggest that urban water and energy efficiency outcomes may be closely tied to borough-specific factors, such as infrastructure age, building stock, and the extent of policy enforcement or retrofitting programs.

Moreover, to explore how energy efficiency varies across different building sectors, a bar chart was created by Tableau using the variable Primary Property Type – Portfolio Manager–Calculated as rows and the average of ENERGY STAR Score as columns. Additionally, null values in the 'ENERGY STAR Score' column were removed to ensure the analysis only reflected valid and measurable performance outcomes. This was particularly important to avoid skewed averages when aggregating scores by category.

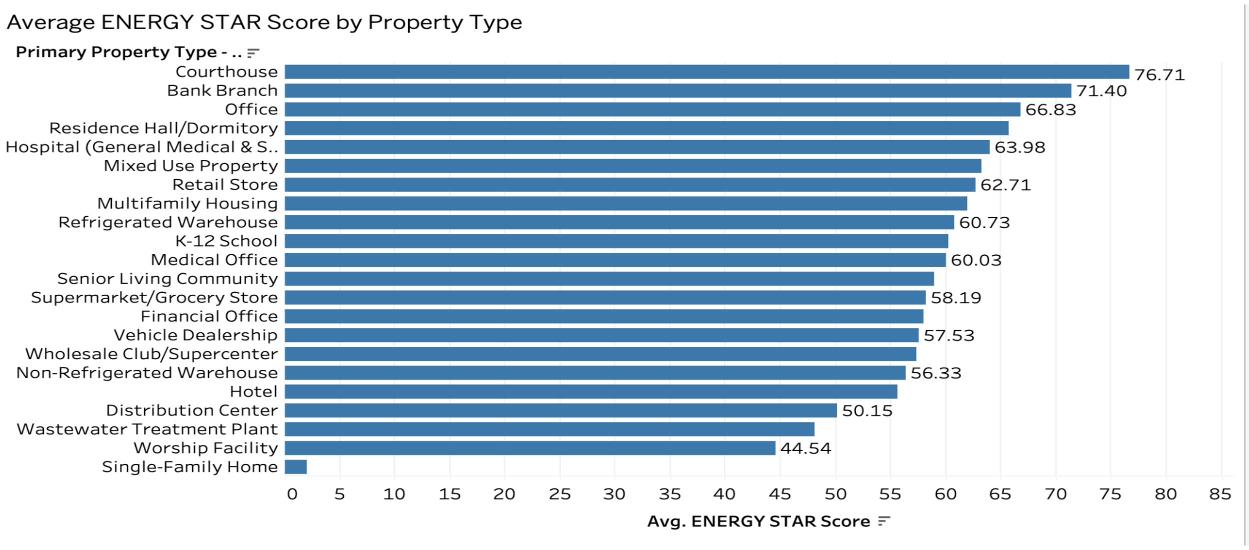

Figure 10. Average ENERGY STAR Score by Property Type

_ Based on Figure 10, Government-related buildings such as courthouses and bank branches achieved the highest average ENERGY STAR Scores (76.71 and 71.40 respectively), indicating strong compliance likely driven by public funding, infrastructure upgrades, and regulatory pressure. Offices, hospitals, and residence halls also scored well (above 63), suggesting policy efforts around institutional buildings may be effective. Multifamily housing and medical offices, common in urban settings, showed moderate scores (around 60–62), reflecting partial adoption of sustainable practices. On the other hand, single-family homes, worship facilities, and wastewater treatment plants recorded the lowest average scores (below 50), suggesting these property types may not be receiving the same level of policy support or may face practical constraints in implementing water and energy-saving improvements.

_ These results suggest that government policies have had uneven influence across sectors. Stronger performance among public, institutional, and commercial buildings may reflect successful enforcement of local laws, such as New York City's Local Law 84, which mandates benchmarking for large buildings. Meanwhile, weaker performance in residential and community-serving facilities points to gaps in coverage or support.

Data Source: Water Related Land Use (Utah Division of Water Resources, 2024)

_ To explore the relationship between water resource management strategies and their impact on agricultural productivity, I used the Water Related Land Use (2023) dataset, which includes 331,643 records for 18 attributes detailing the key contents such as irrigation methods, crop groups, and acreage across Utah. For this analysis, I focused on comparing average acreage managed under each irrigation method.

_ Using Tableau, I created a bar chart that visualises the average of acreage (in acres) per parcel grouped (on y-axis) by irrigation method (on x-axis). This approach allows us to understand the scale at which different irrigation techniques are applied and infer possible relationships between water management practices and land productivity.

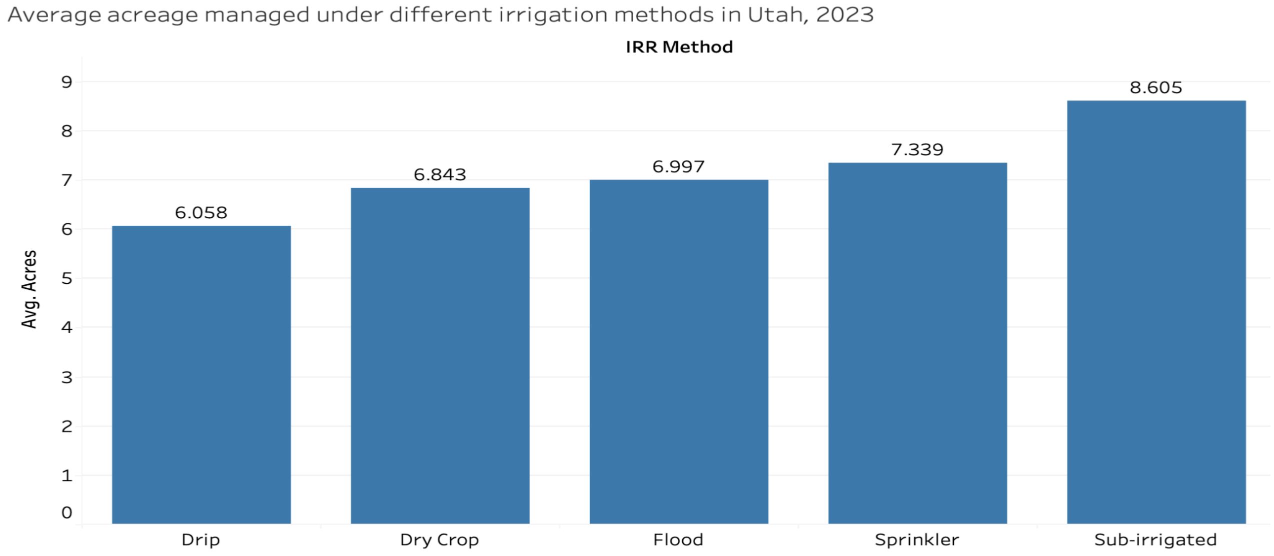

Figure 11. Average acreage managed under different irrigation methods in Utah, 2023.

_ As illustrated in figure 11, the average acreage managed under each irrigation method is relatively close in scale, but some distinctions are still apparent: Sub-irrigated methods had the highest average acreage at 8.605 acres, indicating broader adoption on larger plots, possibly due to its suitability for specific soil types or crop needs. Sprinkler and Flood methods followed with 7.339 acres and 6.997 acres respectively on average, showing their continued relevance in moderate-scale operations. In addition, Dry Crop practices averaged 6.843 acres, suggesting that dryland farming still plays a role in areas with limited water access or minimal irrigation investment. Drip irrigation, although known for its water efficiency, was applied on the smallest average acreage (6.058 acres), which may reflect its higher setup cost and targeted usage in high-value or water-sensitive crop areas.

_ The relatively close range of average acreages across all methods suggests a balanced application of water strategies in Utah agriculture, more efficient systems like Drip irrigation remain underutilized at scale. This could signal opportunities for policy interventions such as subsidies, training programs, or infrastructure support to encourage broader adoption of sustainable irrigation technologies.

To explore the relationship between water resource management strategies and agricultural productivity, I analyzed land use across different water basins in Utah using the cleaned dataset. The bar chart below, created using Tableau, presents the SUM (acreage) on the y-axis managed in each basin on the x-axis for the year 2023.

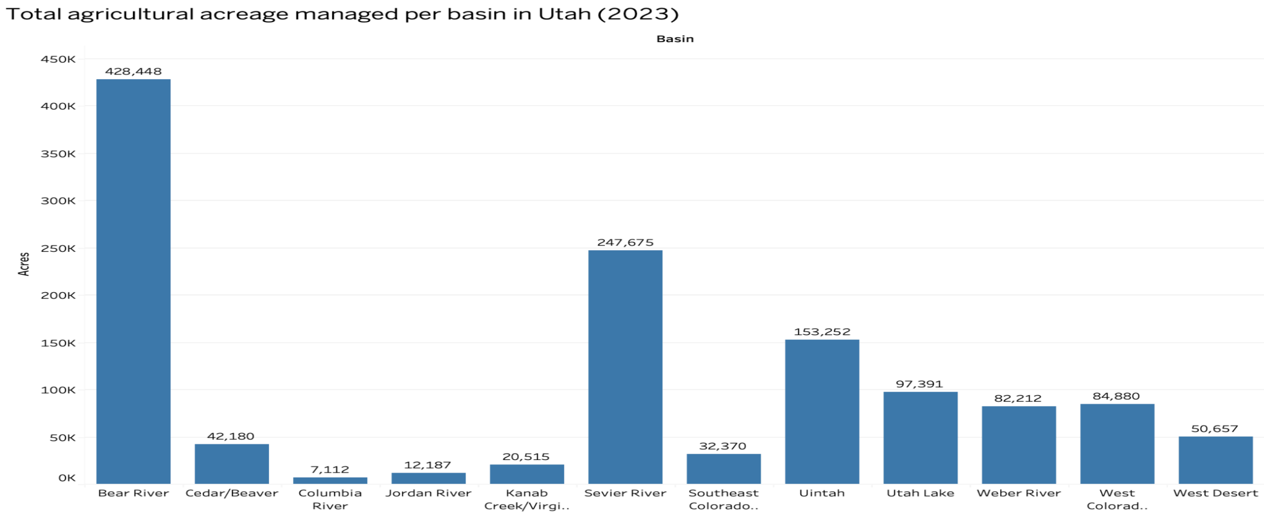

Figure 12. Total Agricultural Acreage by Basin in Utah, 2023

_ Based on Figure 12, the results reveal substantial variation in land use between regions. The Bear River Basin accounts for the largest share of agricultural land, with over 428,448 acres, followed by the Sevier River Basin (247,675 acres) and Uintah Basin (153,252 acres). These basins are likely to be focal points of water demand due to their extensive cultivation areas. In contrast, regions such as the Columbia River Basin (7,112 acres) and Jordan River Basin (12,187 acres) show comparatively limited land use. These differences may reflect a combination of factors, including water availability, infrastructure investment, terrain suitability, and regional policy interventions. From a policy standpoint, the areas with higher acreage suggest a greater dependency on effective water resource management strategies. Without adequate conservation efforts, these basins could face over-extraction and long-term sustainability issues.

Conclusion

Question: What technologies are currently being used in arid regions to enhance water sustainability, and how effective are they?

_ By leveraging data on evapotranspiration (ETo) and precipitation efficiency across various suppliers from Figure 7 and Figure 8, I was able to uncover key insights related to regional water demand and the likely presence of sustainable technologies appear among the top performers in precipitation efficiency, suggesting that their sustainability practices may not be keeping pace with demand. Implementing or enhancing technologies such as weather-based irrigation controllers, stormwater harvesting systems for improved infiltration could help improve water efficiency in these regions. Moreover, Figure 7 provided a clear picture of which areas experience the highest water loss due to evaporation and plant transpiration. Cities like Brawley, Coachella, and Indio reported the highest ETo values, indicating a critical need for advanced irrigation and water conservation measures in these arid zones. In contrast, lower ETo values in places like Los Angeles suggest comparatively less irrigation pressure.

_ To assess how effectively water is being used, I introduced a derived metric: precipitation efficiency. This helped us evaluate how much rainfall is actually retained and utilized for vegetation, although it indirectly reflects the effectiveness of water-saving technologies. Futhermore, most suppliers clustered around an average efficiency of 0.25, but Elsinore Valley Municipal Water District significantly outperformed others with an efficiency of 0.3521 (Figure 8). This suggests the use of advanced techniques such as drip irrigation, smart controllers to enhance water retention and utilization.

_ In conclusion, the data exploration based on Figure 7 and Figure 8 successfully addressed the research question by identifying which regions face the highest water demand and how efficiently different suppliers manage that demand. The findings also highlight areas that may benefit from further investment in sustainable water technologies to ensure long-term resilience in arid environments.

Question: How have government policies in these areas impacted water management practices?

_ From Figure 9, boroughs such as Astoria and Flushing demonstrated relatively high average ENERGY STAR Scores, which may indicate more successful policy enforcement, investment in newer infrastructure, or effective local management practices. In contrast, areas like Jamaica and the Bronx lagged in performance, suggesting challenges such as aging infrastructure or less effective policy implementation. These disparities highlight that location's specific conditions influence how uniformly government mandates are adopted or enforced.

_ At the building-type level, the highest scores were observed in government-related and institutional structures such as courthouses, bank branches, offices, and hospitals (Figure 10). These sectors likely benefit from targeted regulations, public funding, and infrastructure support aligned with the requirements of Local Law 84. Mid-range performers included multifamily housing and medical offices common in urban areas where partial policy uptake may have occurred. Conversely, residential and community-serving facilities like single-family homes and worship centers recorded the lowest average scores, implying these property types might fall outside the main scope of Local Law 84 or face practical barriers to improvement.

_ Overall, the data exploration from Figure 9 and Figure 10, largely answered the research question. It showed that government policies have had measurable, though uneven, influence on energy and water management practices. The clearest evidence of policy impact appears in public and commercial sectors, where mandates such as benchmarking and performance reporting are enforced. Meanwhile, gaps in performance in residential sectors point to areas where policy support or outreach may need to be expanded. This suggests that while Local Law 84 has driven improvements in some domains, broader policy coverage and tailored implementation strategies may be necessary to achieve city-wide sustainability goals.

Question: What are the relationships between water resource management strategies and their impacts on agricultural productivity and urban water supply?

_ Figure 11 presented the average acreage managed under different irrigation methods, and revealed that while most irrigation techniques are applied on similarly sized plots, sub-irrigation had the highest average acreage. This may indicate its broader adoption on larger or more productive fields, potentially due to its suitability for certain soil or crop types. In contrast, drip irrigation, despite its water efficiency, was associated with the smallest average parcel size, suggesting limited large-scale implementation, possibly due to higher upfront costs or specialized application. These patterns point to an opportunity for policy-driven support such as subsidies or education programs to promote wider use of sustainable technologies.

_ Figure 12 examined total agricultural acreage by water basin, highlighting significant variation across regions. Basins like Bear River, Sevier River, and Uintah accounted for the majority of agricultural land, suggesting they are key zones for water demand and management intervention. Meanwhile, basins such as the Columbia River and Jordan River showed much smaller cultivated areas. These disparities reflect the combined effects of physical terrain, water availability, infrastructure, and local policy.

_ Together, the visualisations (Figure 11 and Figure 12) provide strong insight into how water strategies and policy influence land use outcomes. While the dataset does not explicitly include policy variables, patterns in irrigation scale and regional distribution suggest indirect evidence of strategic water management efforts. The findings support the hypothesis that sustainable irrigation practices and regional planning directly impact agricultural productivity, and that further policy intervention could enhance the scalability of efficient water practices across Utah.

Reflection

_ This project offered valuable hands-on experience in data cleaning, exploration, and visualisation using real-world datasets on water resource management and energy efficiency. One of the key lessons that I learned was the importance of thoroughly cleaning and preprocessing data before performing analysis. I initially encountered inconsistencies in the formatting of city names within the NYC Building Energy and Water Data Disclosure dataset. This experience reinforced the importance of cross-verifying data across multiple tools, such as Python and Tableau, and applying flexible techniques to identify and handle inconsistent attribute formats during the cleaning process.

_ Another takeaway was the importance of aligning visualisations with specific research questions. Initially, I generated a number of exploratory plots without a clear analytical direction. However, refining my focus to match each visualisation with a targeted research question allowed me to derive more meaningful and actionable insights.
_ Overall, this project highlighted the value of an iterative, question-driven approach to data science and underscored the importance of balancing technical analysis with domain-relevant interpretation.

Bibliography

Liu, X, Tang, Q., Zhao, Y., & Wang, P. (2024). Persistent water scarcity due to high irrigation demand in arid China: A case study in the North Slope of the Tianshan Mountains. Earth's Future. Retrieved from: https://agupubs.onlinelibrary.wiley.com/doi/full/10.1029/2024EF005070

Zhang, L., Yu, Y., Guo, Z., Ding, X., Zhang, J., & Yu, R. (2024). Investigating agricultural water sustainability in arid regions with Bayesian network and water footprint theories. Science of The Total Environment, 951, 175544. Retrieved from: https://www.sciencedirect.com/science/article/pii/S0048969724056948

California Department of Water Resources. (2023, December 6). Fiscal Year Annual Data – CSV. DataSetsAI. Retrieved from: https://datasets.ai/datasets/data-for-calculating-efficient-outdoor-water-uses-147dd

NYC Mayor’s Office of Sustainability. (2022). NYC Building Energy and Water Data Disclosure for Local Law 84 (2022–Present). NYC OpenData. Retrieved from: https://data.cityofnewyork.us/Environment/NYC-Building-Energy-and-Water-Data-Disclosure-for-/5zyy-y8am/data_preview

Utah Division of Water Resources. (2024, March 21). Water Related Land Use (2023). Utah DNR Online Maps. Retrieved from: https://dwre-utahdnr.opendata.arcgis.com/datasets/water-related-land-use-2023/explore?location=39.581924%2C-111.604167%2C6.70

Sustainability Directory. (2025, March 31). Precipitation Efficiency. Retrieved from: https://climate.sustainability-directory.com/term/precipitation-efficiency/

Zeng, X., & Rouse, K. H. (2022, December 5). Using bar charts to compare data in categories. Journal of AHIMA.

Retrieved from: https://journal.ahima.org/page/using-bar-charts-to-compare-data-in-categories

Magdy. (2024, October 30). Sustainable resource management: A pathway to a greener future – A guide to building resilient systems for future generations. IEREK. Retrieved from:

https://www.ierek.com/news/sustainable-resource-management-a-pathway-to-a-greener-future/