**Dylan Johnson  
US Water Crisis: Are We Running Out of Water?  
DATA 205 | CRN 34669**

**Introduction & Project Overview**

Water scarcity is becoming an increasingly urgent issue in the United States, particularly in regions heavily dependent on limited water resources. This project aims to investigate national water use trends, project future water demand, and evaluate sustainability against thresholds based on historical maximum consumption. The research brings together multiple datasets representing public water supply, thermoelectric water use, irrigation, population trends, and water quality metrics such as salinity and pH. This analysis is critical for understanding when the U.S. may exceed its sustainable freshwater capacity and to inform policy and infrastructure decisions that can mitigate long-term risk.

**Data**

**The analysis incorporates data from five major sources:**

* Water\_Use\_10States.csv: Public supply water usage across ten U.S. states, in million gallons per day (MGD).
* TEWU\_10States.csv: Thermoelectric water use by HUC regions and year.
* Irrigation.xlsx: Agricultural irrigation data, including state-level MGD usage.
* Population\_data.csv: Demographic projections including total population, broken down by sex, race, and origin.
* BKB\_WaterQualityData\_2020084.csv: Water quality monitoring data with metrics such as salinity and pH over time.

**Excluded data**: Some water quality entries lacking date or site identifiers, and incomplete population records for subgroups, were excluded during cleaning.

**Goals**

The primary goal is to determine when the U.S. will surpass its sustainable water use threshold, by combining historical and projected data across key sectors. This effort will help identify whether current consumption patterns are compatible with long-term water availability.

**Tools**

All data processing and analysis were performed in R, leveraging packages such as tidyverse, ggplot2, janitor, readxl, and cluster. Linear regression was used for forecasting, and ggplot2 was used for visualizations. Final graphs were exported and included to support analysis.

**Data Cleaning and Pre-processing**

* Water use datasets: Converted from wide to long format using custom tibble mappings and pivot\_longer(). Missing years and invalid MGD entries were removed.
* TEWU and Irrigation: Reformatted using parallel transformations with bind\_cols() and variable alignment.
* Population: Aggregated into a total\_pop variable, with sex, race, and origin fields recoded for clarity. A polynomial model extrapolated data back to 2000.
* Water quality: Filtered for missing values and classified pH and salinity using EPA guidelines.

**All pre-processing steps are documented in the GitHub repository.**

**Descriptive Statistics**

* Public Supply: 0.5126 MGD (mean)
* Thermoelectric: 0.0359 MGD (mean)
* Irrigation: 0.5126 MGD (mean)
* Population Projections: From ~282 million (2000) to a peak of ~740 million (2080), then slightly declining.
* Water Quality: Most pH values are safe (6.5–8.5), though high salinity (>0.5 ppt) appears sporadically.

**Main Results and Interpretation**

**Figure 1: Projected US Water Demand vs. Sustainability Threshold (2000–2100)**

This graph shows projected total water demand from 2000 to 2100, generated by a linear model fit on historical use. The red dashed line indicates the sustainability threshold — the historical peak usage.

**Key Finding:** U.S. water demand exceeded the threshold in 2018 and is projected to rise to over 1.2 trillion gallons per year by 2100.

**Figure 2: Sectoral Contributions to Water Demand**

A stacked area chart illustrates the cumulative impact of the public, thermoelectric, and irrigation sectors. The public and thermoelectric sectors dominate total use.

**Insight:** Conservation strategies should prioritize these two sectors for maximum impact.

**Figure 3: Water Use vs. Salinity Trends**

Salinity levels vary independently of water use volume, indicating external factors such as runoff, coastal intrusion, or industrial waste may play a larger role in water quality variation.

**Additional Analyses and Visualizations**

**Figure 4: Per Capita Water Use (GPCD) Forecast**

A new projection calculates daily per-person water usage through 2100. The data reveals that per capita water use will continue increasing, driven by both rising demand and population.

**Figure 5: Clustered States by Water Use**

States were grouped into clusters using k-means clustering based on average public water usage. This helps identify regional similarities in usage intensity and prioritize region-specific interventions.

**Figure 6: State-Level Water Efficiency Trends**

Gallons per capita per day (GPCD) were calculated for each state by distributing national population using proportional water use. This revealed significant variation in efficiency — some states deliver higher water volumes per person due to infrastructure, climate, or policy.

**Implications and Recommendations**

* Infrastructure Planning: Public utilities should prepare for increased demand, particularly in states with high per capita use.
* Water Quality Monitoring: Salinity issues in some regions require additional investigation and investment in treatment or alternative sourcing.
* Sustainability Planning: The threshold was crossed in 2018. Future policy must address projected overshoot through conservation, desalination, or improved technology.
* Data-Driven Interventions: States clustered with high water use may benefit from targeted conservation campaigns, leak mitigation, or incentive programs for efficiency.

**References and Acknowledgements**

**References:**

* U.S. Geological Survey datasets
* EPA Water Quality Standards
* Census Bureau Demographic Projections

**Acknowledgements:**

* Prof. Perine for project mentorship
* Open-source R contributors and package maintainers
* The GitHub community for collaborative tools

**GitHub Repository:**[**https://github.com/dylanj141/Data205---Dylan-Johnson**](https://github.com/dylanj141/Data205---Dylan-Johnson)