Project 6

Beau Martin, Joshua Christensen

Introduction

In this project, we explore the field of hydrology, focusing on the analysis of baseflow patterns in the Republican River Basin using multiple linear regression. Baseflow, the groundwater contribution to river flow, is crucial for understanding water availability and sustainability in regions dependent on both natural flow and human-driven water use. Using linear regression, we built a predictive model that can estimate baseflow based on these influential factors. With this model we are able to see the relationships between baseflow and factors such as precipitation, Irrigation, and evapotranspiration. Here is our code and slides.

Dataset

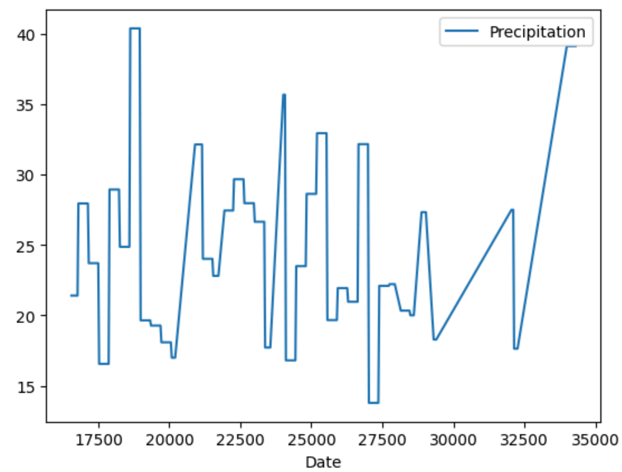
For our analysis, we utilize a hydrologic dataset with monthly observed baseflow data from 42 segments within the Republican River Basin. This dataset is particularly suited to our domain as it includes critical attributes—such as precipitation, evapotranspiration, and groundwater extraction for irrigation—that are integral to understanding water balance and resource management in river basins. Together, these features provide a comprehensive view of hydrologic processes impacting water availability and usage within the basin.

Analysis Technique

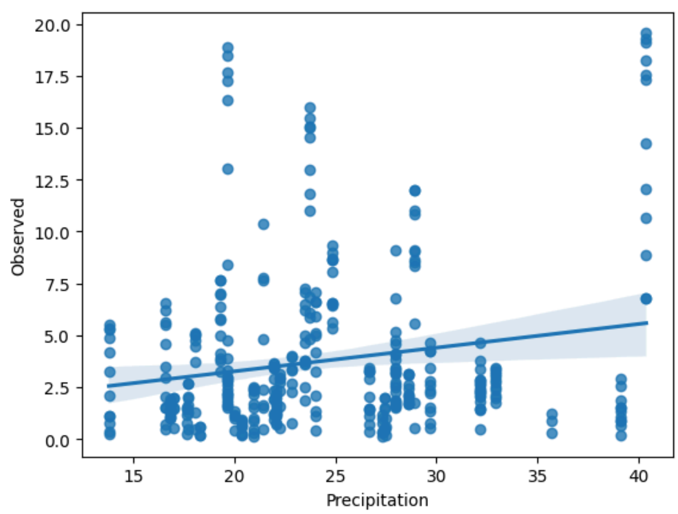
To help visualize our data and relationships, we used scatter plots and line graphs. To investigate more into precipitation's effect on baseflow we found the segment that had the highest ever recorded rainfall and removed all records that had a precipitation of 0. Next, we created a linear regression plot to help visualize our relationship between precipitation and baseflow. To evaluate the relationship, we performed a t-test on the model coefficients.

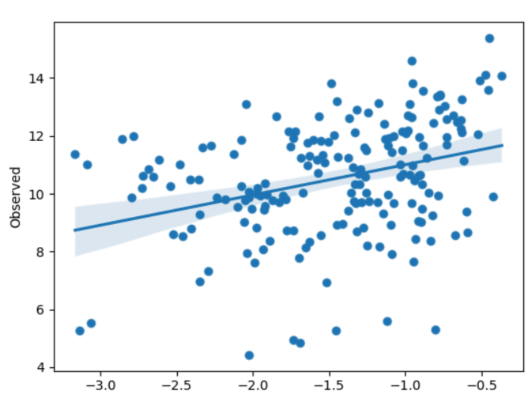
Results

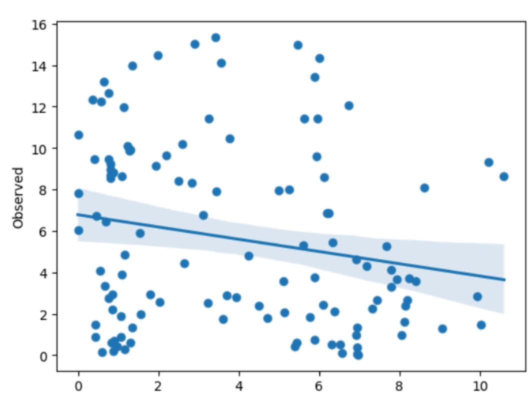
To examine how precipitation changes over time, we created a scatter plot showing annual precipitation totals for each year in segment 171. This visualization allowed us to observe patterns and variations in precipitation levels across different years. We then observed two other values in different segments, both selected for the widest variation for the chosen statistics. Comparing across statistics, it became clear that Precipitation increased baseflow, while Evapotranspiration and Irrigation Pumping generally leads to a reduction in baseflow.



Next, we developed a linear regression plot to explore the relationship between precipitation and observed baseflow. This plot includes the least-squares regression line and a 95% confidence interval, providing a clear view of the predictive relationship. To quantitatively assess this relationship, we performed a t-test on the model coefficients, resulting in a p-value of 0.000348. This low p-value indicates that precipitation is a statistically significant predictor of observed baseflow.







Technical

Before we were able to perform our linear regression, we made our date column more manageable by subtracting 693963 days to make our days start at Jan 1, 1900 instead of Jan 1, 0000. For our precipitation model we wanted to look at a segment that had the highest recorded rainfall which was segment 171. Then, we removed all records that contained a precipitation of 0 so that we are only looking at records that show the relationship of how the amount of rainfall affects our baseflow. We created our regression plot for our relationship using seaborn and then we performed our t-test on the model coefficients using scikit-learn. We then performed truncated versions of the same analysis on Irrigation and evapotranspiration in segments with the most extreme (positive or negative, depending) weight for the selected variable.