**Step 1: Pull data from USGS website**

**Code name: Step1\_PullDataFromUSGS**

1. Use the shapefile to pull data from the USGS website

desoto\_shapefile\_path <- "/Users/ahowl/Desktop/KGS Data analysis/DeSoto\_shp/DeSoto.shp"

watershed\_shapefile\_path <- "/Users/ahowl/Desktop/KGS Data analysis/WatershedBoundary\_KN\_20230113/watershed\_bndry.shp"

2. Map created with station data (mean streamflow, parameter cd=X\_00060\_00003

parameterCd = "00060": Requests streamflow data in cubic feet per second.

X\_00060\_00003: Represents the mean daily streamflow in cubic feet per second (cfs)

3. a tibble with cleaned up data

4. Check pulled up data from code to usgs station

**Output:**

1. Step1\_Output1\_WebMapper\_meanstreamflow.html

(file with interactive mapper for the stations)

2. streamflow\_tibbles\_step1.rds

(all the preliminary streamflow data)

**Step 2: Streamflow data qaqc**

**Code name:** Step2\_StreamflowData\_QAQC

1. Found the sampling years of the stations (plots: timeline)

2. Found missing data

3. Found stations with less than 10 percent missing data

Output:

1. Step2\_timeline figs (shows years when the stations were sampled)

2. streamflow\_tibbles\_Filtered\_step2 (filtered tibble with less than 10 percent missing data)

3. Histogram of a) percent of missing data vs stations b)num of years vs. stations

**Step2a: Converting data into MMD**

**Code name: Step2a\_watershedAreaCalculation\_MMD**

**1. Data Preparation:**

* Calculates watershed area for each station and adds it to the data frame.
* Identifies stations with missing area data.

**2. Unit Conversion:**

* Converts streamflow data from CFS to MMD using watershed area.
* Adds a new column mean\_streamflow\_mm\_per\_day containing the MMD values.

**3. Data Filtering and Visualization:**

* Filters out canal stations and creates separate data frames.
* Plots stations with missing area data on a map.

**4. Output:**

* Saves the data frame with MMD values and the list of stations with missing area data to RDS files.
* Saves individual plots comparing CFS and MMD for each station.

**Step 3: Streamflow data separating Medium and Long term subset**

**Code name:** Step3\_SeperatingLongMediumSubset

1. Medium term: data on >90% of days between 1/1/1979 and 12/31/2023

Long term: data on >90% of days between 1/1/1944 and 12/31/2023

2. Made a dataframe from start and end date and combine it with the original data to find out the missing values

3. Found stations with less than 10 percent missing data

4. Timeline figures

5. Stations figures

Output:

1. Step3\_timeline figs (shows years when the stations were sampled for Medium and Long term)

2. streamflow\_tibbles\_Filtered\_LongSubset\_step3.rds

3. streamflow\_tibbles\_Filtered\_MediumSubset\_step3.rds

4. Step3\_Streamflowdata\_locations\_LongSubset (Location map of Long term subset)

5.Step3\_Streamflowdata\_locations\_MediumSubset (Location map of Medium term subset)

**Step 3: Streamflow data annual average, seasonal average (of the years)**

**Code name:** Step4\_AnnualAverage

1. Annual and Seasonal Calculation: Calculates the annual and seasonal mean streamflow for each site using summarise() and groups by year and season.

2. Combining Data: Merges the annual and seasonal streamflow data into a single dataset (Combined\_Data).

3. Subset Creation: Filters the AllYear\_StreamflowData\_Annual dataset for medium-term and long-term subsets based on site numbers

Output:

1. LongTerm\_StreamflowData\_Annual

2. MediumTerm\_StreamflowData\_Annual

3. AllYear\_StreamflowData\_Annual