Proposal and Analysis Plan

Rachel Bash, Caroline Watson, Keqi He, Haoyu Zhang

October 25, 2019

What questions will your team address?

1. What is the current state of the rivers in the Southeastern Missouri River Basin and how has it changed over time?
2. What effects did the March 2019 floods have on the water quality and quantity of rivers in the Missouri River Basin areas of interest?
3. What effects did the 2012-2013 drought have on the water quality and quantity of rivers in the Missouri River Basin areas of interest?
4. Given past and current data, what can we predict about the future state of water in the Missouri River Basin?

What hypotheses will your team address?

- Hypotheses should relate directly to your questions. Each numbered hypothesis should match up with the correponding numbered question.
- There may be multiple working hypotheses for a single question. If this is the case, note each hypothesis as 1a, 1b. etc.

1. Water quantity and quality in the Missouri River Basin has become more variable (more frequent storms and droughts) over time.
2. Nutrient concentrations and flow volume increased due to the floods of the 2019 flood.(we will create a hysteresis plot and a hydrograph for each site and nutrient of interest)

a. Nitrate concentration will increase with flow. We hypothesize that hysteresis plots will exhibit a clockwise direction due to flushing of nutrients from the quickflow.

3. Discharge will decrease during a drought, also causing a decreased concentration of nitrogen due to less overland flow.
4. We predict that total flow in the Missouri River Basin is decreasing (non-stationary), and so the future situation of the river basin will see the continuation of current trends of decreasing overall volume of flow.

What dataset(s) will your team analyze?

USGS, Water Quality Portal

How will you set up and manage your project repository?

We have set up a git repository, so we will pull and push from the same master.

Create a table of variables you will analyze.

- Column 1: Variable
- Column 2: Units (if known)
- Column 3: Dependent (response) or independent (predictor) variable?
- Column 4: To which hypothesis(es) does this variable pertain?

Note: You may not know all of the individual variables you plan to analyze at this point. It is sufficient to describe what type of variable you anticipate using (e.g., land cover) and decide on specifics later

```
library(tidyverse)

Variable <- c("discharge", "time", "nitrogen", "pH", "total coliform",
              "O2 Concentration", "Phosphorus")

Units <- c("cfs or m3/s", "UTC", "mg/L", "1", "cfu/100mL", "mg/L", "mg/L")

TypeOfVariable <- c("both", "independent", "dependent", "dependent", "dependent",
                    "dependent", "dependent")

Hypothesis <- c("all", "all", "2 and 3", "2 and 3", "2", "2 and 3", "2 and 3")

Table <- cbind(Variable, Units, TypeOfVariable, Hypothesis)

knitr::kable(Table)
```

Variable	Units	TypeOfVariable	Hypothesis
discharge	cfs or m3/s	both	all
time	UTC	independent	all
nitrogen	mg/L	dependent	2 and 3
pH	1	dependent	2 and 3
total coliform	cfu/100mL	dependent	2
O2 Concentration	mg/L	dependent	2 and 3
Phosphorus	mg/L	dependent	2 and 3

What tasks will your team conduct?

Data acquisition

We will get our data from USGS (NWIS Site) and the Water Quality Portal by looking for sites with good available data in the Missouri Region (Hydrologic Region 10). We will pick 2-4 sites from each HUC 4 in the 1020-1029. We will choose sites that are both on the Missouri River and sites that are on tributaries of the Missouri River. We will only pull USGS sites on rivers.

We will clean our dataset by getting rid of unnecessary columns to make analysis easier and will rename columns for clarity and conciseness.

We will also pull the watershed boundary dataset from USGS to map our sites and water features.

Data exploration

Visualizing our site locations and basic discharge behavior will be our first steps in exploring our data. We will visualize basic discharge behavior and water quality data.

After exploration, we will address NA's if they are affecting our data.

Data wrangling

We will create subset tables in order to focus on specific sites when we address our flooding and drought questions. We will group our data by year when we want to address trends over time. We will also wrangle data by summarizing by month (and by day) from multiple years to get average, maximum, and minimum values to compare. In order to find the recurrence interval, we will need to wrangle data in order to find peak discharge.

Data analysis and visualization

We will create hysteresis plots, recurrence interval plots, discharge over time, typical discharge pattern over a year (plot with min, max, and mean), nutrient concentration patterns over a year (specifically for 2019), pH over time and over flow, total cell count over time, total coliform over time, plot hydrograph separation between baseflow and quickflow. We will be creating dygraphs to visualize time series for discharge and, if applicable, nutrient and oxygen concentrations.