

# EXPLORING THE RELATIONSHIP BETWEEN STREAMFLOW STATISTICS AND DRAINAGE AREA

Group 2: Meelisha Maharjan, Xinchun He

# Background

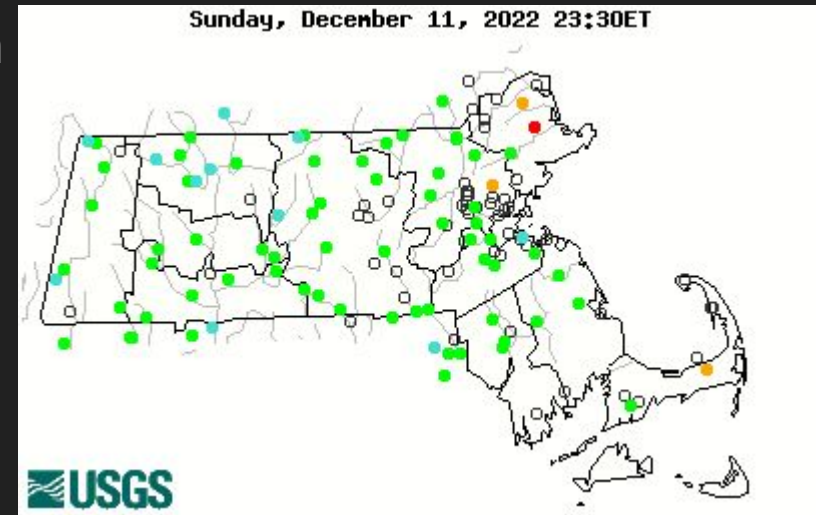
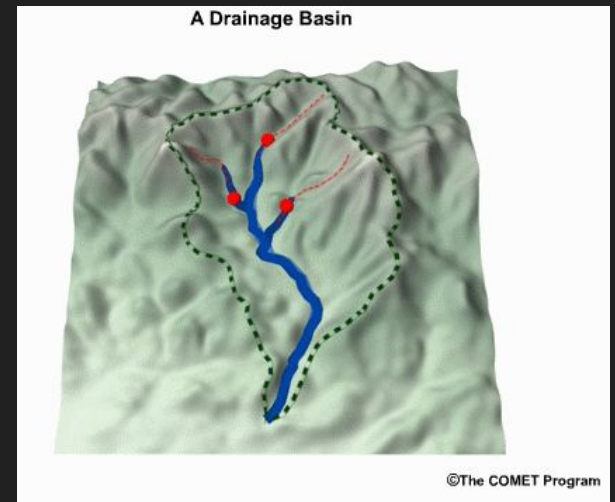
- Drainage area is a major factor in streamflow generation
- Ungauged rivers

## Objective:

- Establish a statistical relationship between the drainage area  $\sim$  Q5, Q95
- Q5: Highflow; Q95 Lowflow

## Study region:

- Massachusetts region : 149 gauges in Massachusetts



# Methodologies

## Data scraping:

- Using regex to scrape 20 year daily streamflow data from USGS website

## Data cleaning:

- Filtered out unqualified data (< 15yrs and NAs)

## Statistics computation:

- Calculating streamflow statistics (quantiles)
- Preview drainage area VS Q95/Q5

The screenshot shows the USGS National Water Information System Web Interface. The page is titled "USGS 08324000 JEMEZ RIVER NEAR JEMEZ, NM" and "PROVISIONAL DATA SUBJECT TO REVISION". It includes a "Summary of all available data for this site" and an "Instantaneous-data availability statement". The table below shows the data available for this site.

Date / Time	Discharge, (CFS)	Gage height, feet
12/06/2022 00:00 MST	31.0 <sup>0</sup>	3.60 <sup>0</sup>
12/06/2022 00:15 MST	31.0 <sup>0</sup>	3.60 <sup>0</sup>
12/06/2022 00:30 MST	31.0 <sup>0</sup>	3.60 <sup>0</sup>
12/06/2022 00:45 MST	31.0 <sup>0</sup>	3.60 <sup>0</sup>
12/06/2022 01:00 MST	31.0 <sup>0</sup>	3.60 <sup>0</sup>
12/06/2022 01:15 MST	31.0 <sup>0</sup>	3.60 <sup>0</sup>

# Methodologies

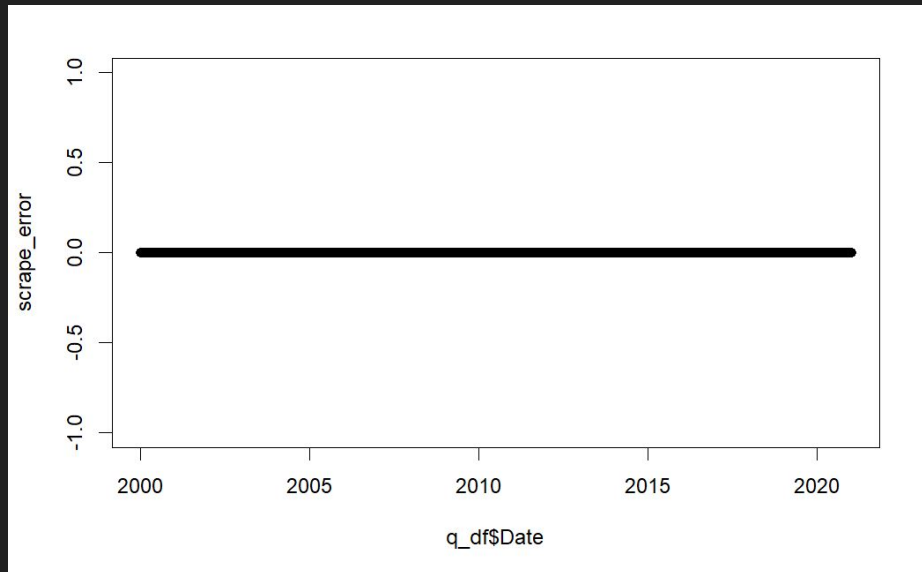
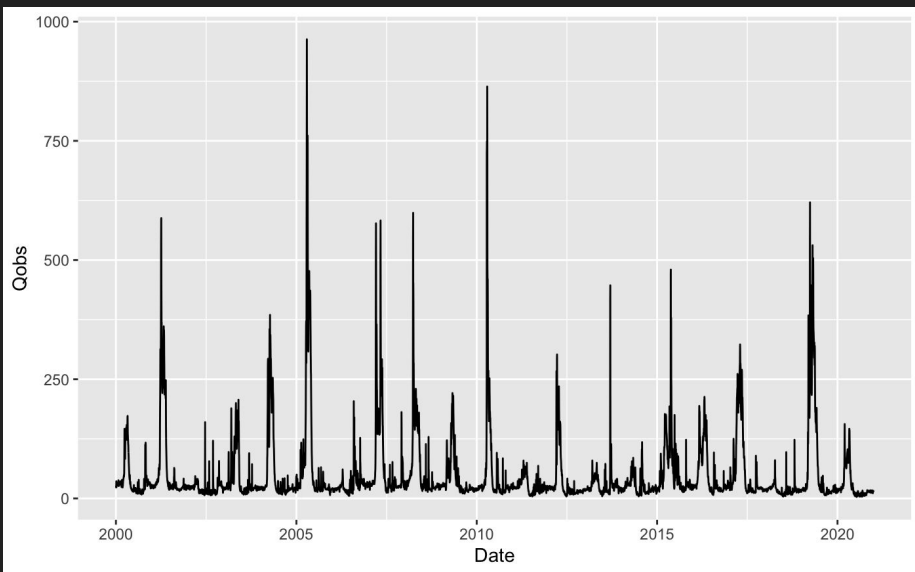
## **Model fitting**

Optimization Function: `Optim`

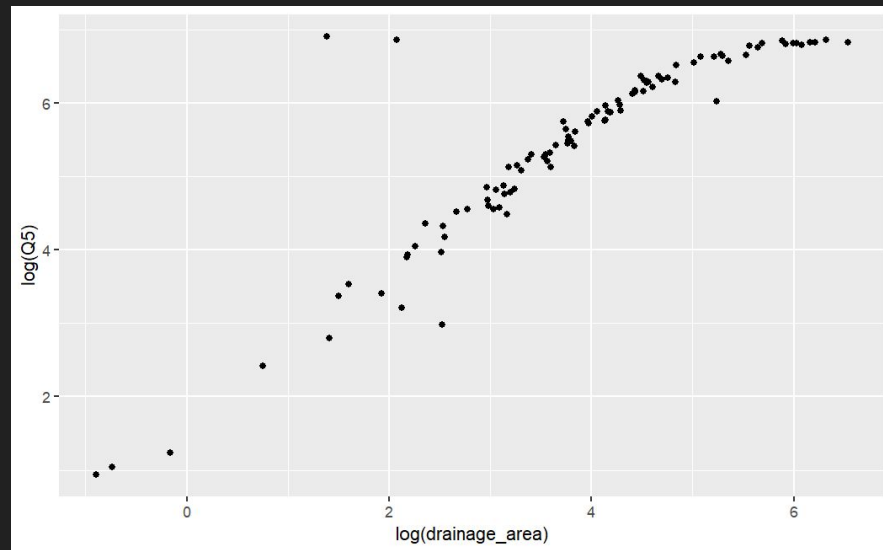
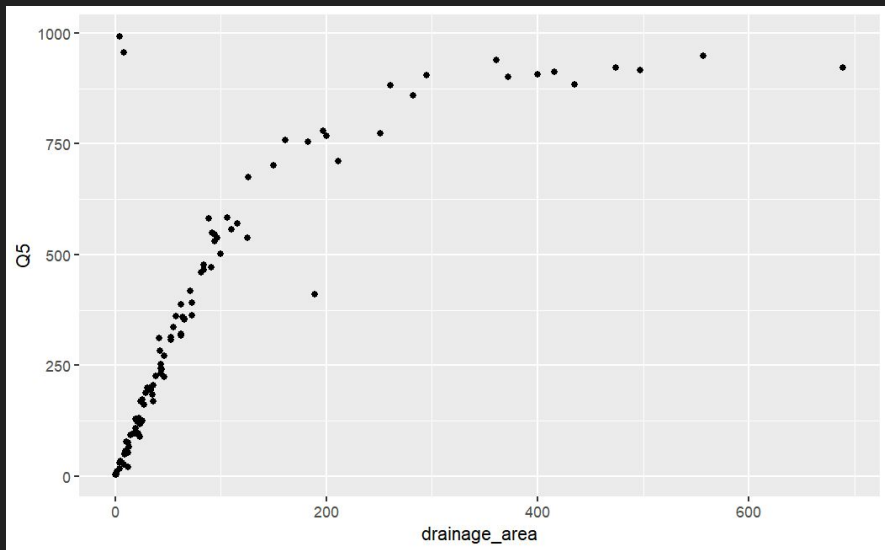
Bootstrapping Method: robustness; confirm relationship between variables

# Data scraping validation plot (`dataRetrieval`)

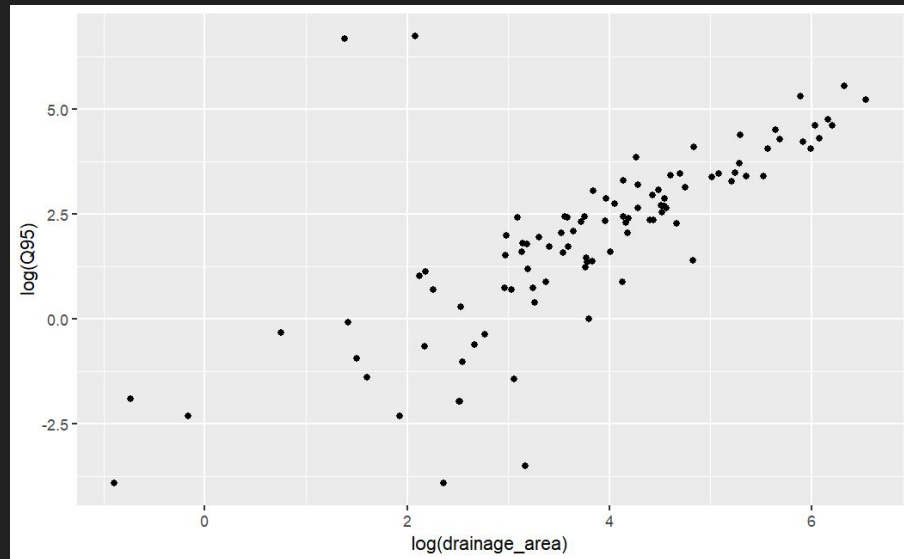
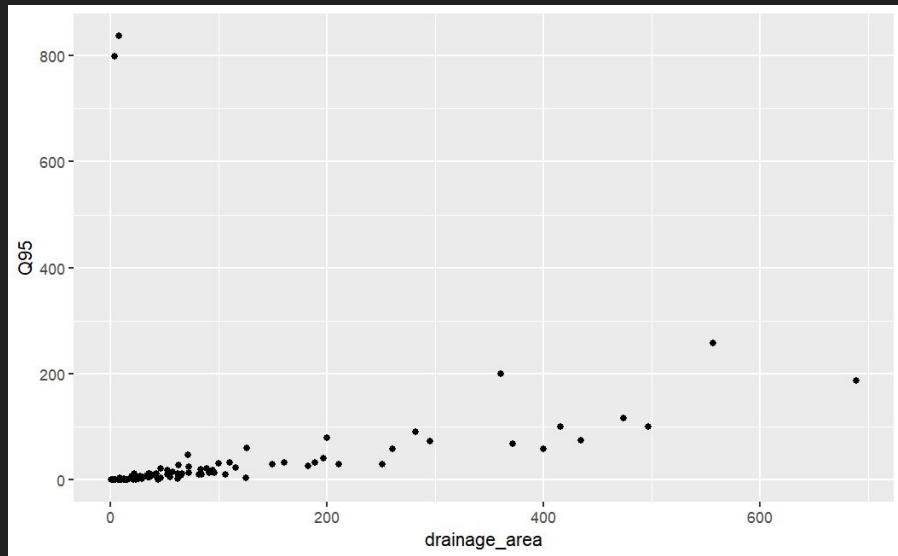
USGS 08324000 JEMEZ RIVER NEAR JEMEZ, NM



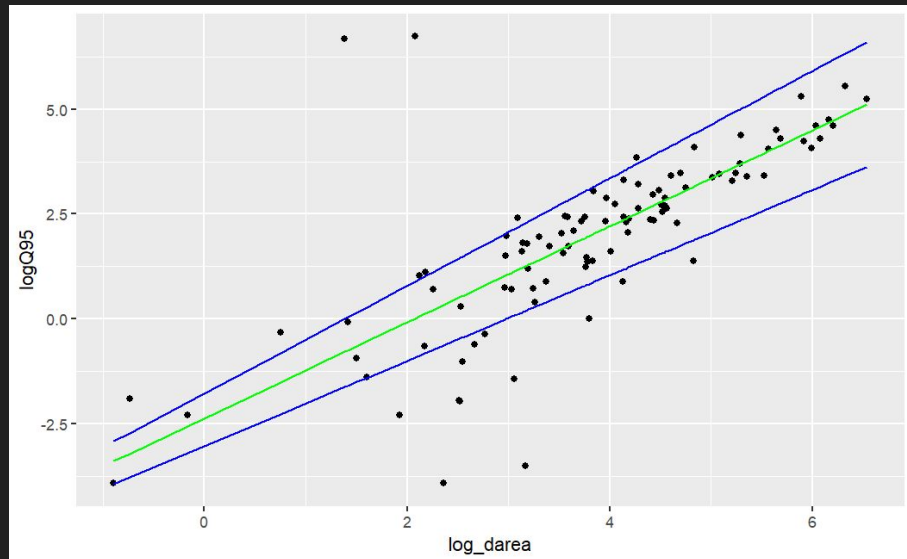
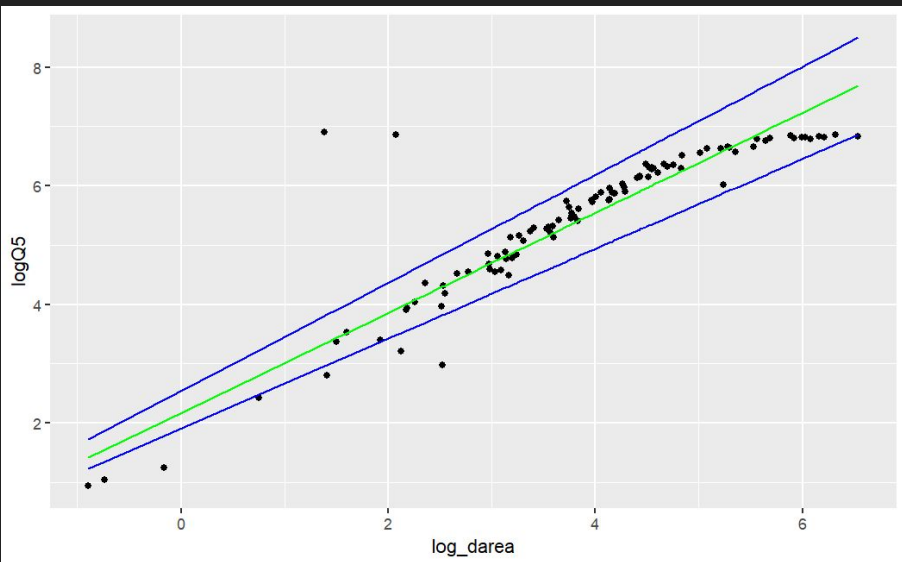
# Drainage Area VS Q5: normal & log



# Drainage Area VS Q95: normal & log

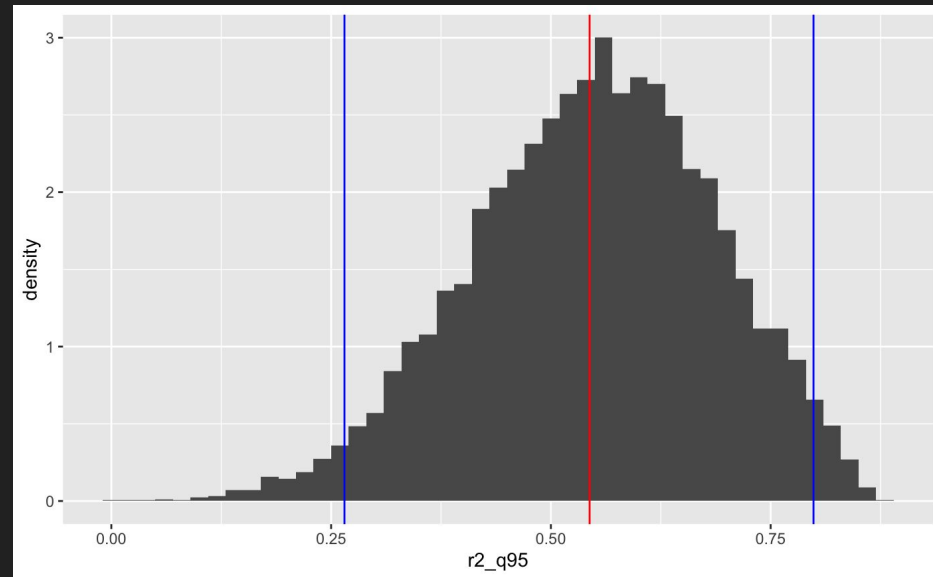
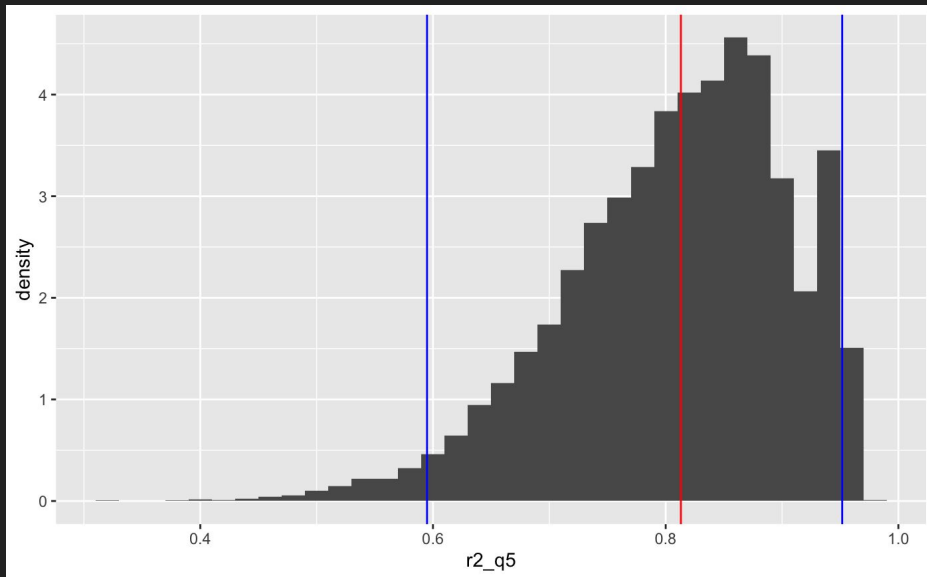


# Ensemble regressions from the bootstrap sampling

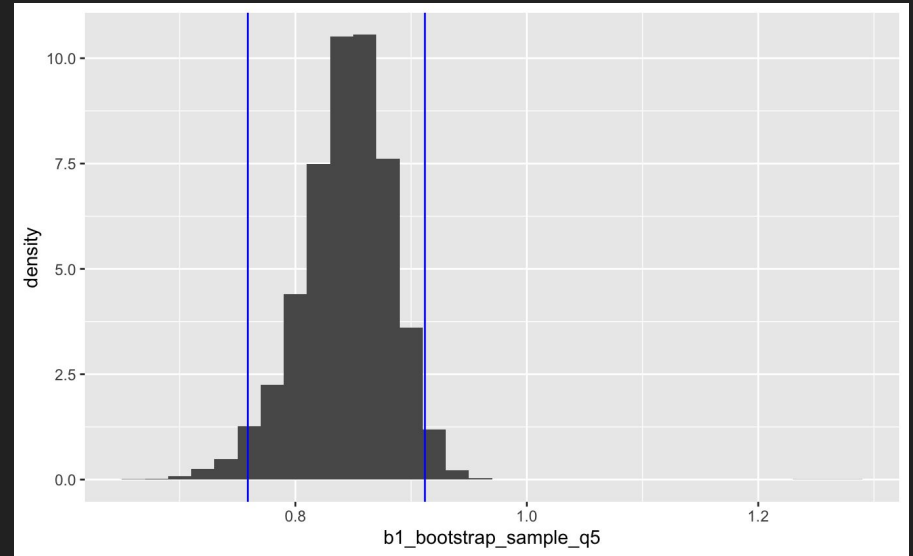
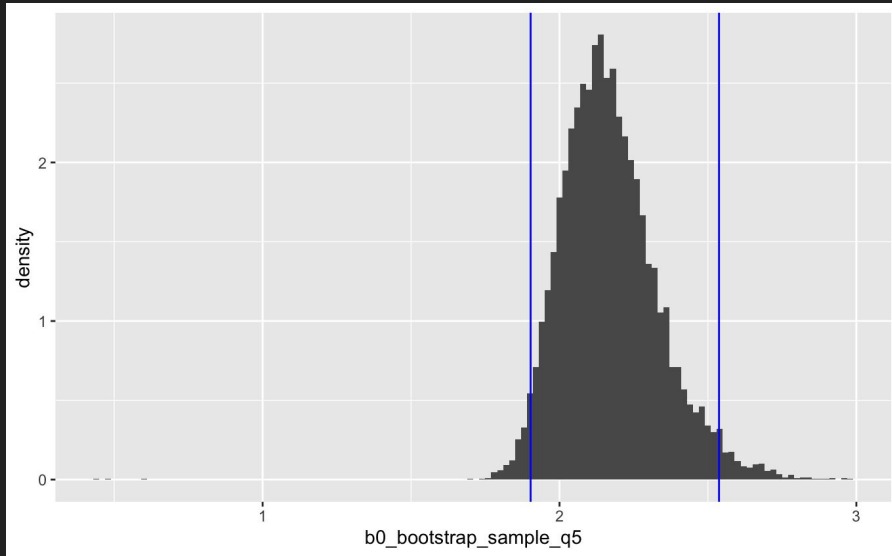




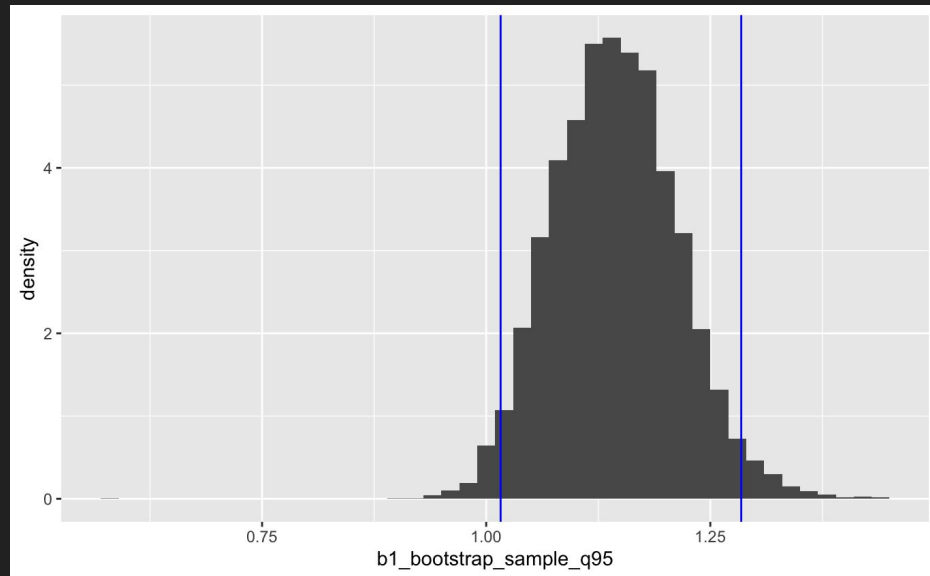
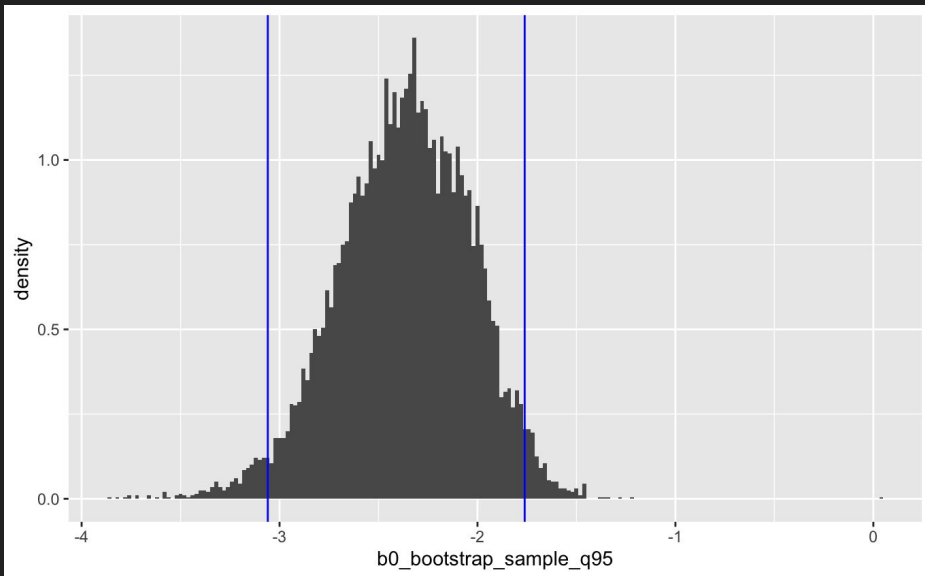
# R-squares



$$\log(Q5) = b0 + b1 \times \log(\text{Drainage Area})$$



$$\log(Q95) = b0 + b1 \times \log(\text{Drainage Area})$$



# Conclusions

- Automatic web-scraping tool to obtain streamflow/drainage area data from USGS gauges
- Using bootstrap sampling method to build a regression model between drainage area and  $Q_5/Q_{95}$



# Background

- Streamflow: Streamflow (also known as discharge) is the volume of water flowing past a given point in the stream in a given period of time. Streamflow is reported as cubic feet per second (ft<sup>3</sup>/s). [USGS]
- Q5: High flows of a streamflow that is related to 5% or more exceedance probability.
- Q97.5 : These are the low flows of streamflow that have 97.5% or more exceedance probability.
- Drainage Area: The term "drainage area" is defined as the land area where precipitation falls off into creeks, streams, rivers, lakes, and reservoirs. It is a land feature that can be identified by tracing a line along the highest elevation between two areas on a map, often a ridge. [USGS]
- Streamgage: A streamgage is a structure installed beside a stream or river that contains equipment that measures and records the water level (called gage height or stage) of the stream.