An Analysis of Extreme Hydrologic Trends in Dallas, Texas

https://github.com/jakegreif/EDA_Final_Project

Jake Greif

Abstract

Across the country, the expected effects of climate change are being realized in the form of more intense precipitation events and dry periods. In North Texas, the city of Dallas is familiar with both types of extreme hydrologic events, and it has experienced particularly intense events in recent years. An analysis was conducted to determine if floods and droughts, defined as the top and bottom quartiles of daily average discharge, repsectively, have become more frequent in the Trinity River and a tributary near Dallas, TX. The results show that drought discharges have decreased, and flood discharges have slightly increased over time in the Dallas area, and that extreme discharge trends in a larger, more regulated river like the Trinity River are less intense than a lower order tributary.

Contents

List of Tables

List of Figures

1 Research Question and Rationale

Across the country, the expected effects of climate change are being realized in the form of more intense precipitation events and dry periods. In North Texas, the city of Dallas is familiar with intense storms as it resides in a portion of the country where warm air from the Gulf of Mexico and cool air flowing down from the Rocky Mountains converge, producing dramatic storms. Though, the city is also accustomed to dry spells, particularly in the summer. Despite the famaliarity this region has with both hydrologic extremes, it has experienced particularly intense events in recent years. Therefore, this project will examine trends in the flow of the Trinity River and a tributary of the Trinity River near Dallas, TX. In particular, the analysis will attempt to determine if extreme flow rates have become more prevalent in recent decades compared to discharges over the past centry.

Have extreme discharge events become more frequent in the Dallas area? How do the frequency of extreme discharge events in higher and lower order streams in the area compare? To answer this question, data from both USGS site 08057000 on the Trinity river in Dallas, TX and USGS site 08049500 on the West Fork Trinity River in Grand Prairie, TX will be analyzed. Specifically, the mean daily discharge at both sites will be characterized based on the frequency of occurance. As per USGS percentile thresholds, a discharge that falls above the 75th percentile is considered to be above normal, and a discharge that falls below the 25th percentile is considered below normal. Therefore, in the context of this analysis, a discharge above the 75th percentile is considered a "flood", and a discharge below the 25th percentile is considered a "drought." The goal of this analysis is to determine if recent extreme events are indicative of an emerging trend, as well as if a larger, more regulated river (i.e.the Trinity River) has less frequent extreme discharge than a lower order tributary.

2 Dataset Information

The dataset used in the analysis was sourced from the U.S. Geological Survey's (USGS) Water Data website. While there is a collection of physical and chemical data available, the only data used for this analysis is average daily discharge (ft^3/s). The gauge sites used in the analysis were found using an interactive map on the USGS Water Data website, and were selected based on their proximity to the city of Dallas. The West Fork Trinity River was chosen as the representative tributary due the lack of development surrounding the river, as well as the lack of river control present upstream. After idenifying the gauge station number at both locations, the number was entered into the 'Site Number' search bar. Then, the 'Discharge' check-box was selected and the years 1925-2019 were selected. 1925 was chosen as the start date based on the earliest available data for both datasets. Both datasets also contain maximum and minimum daily discharge, however that data is only available beginning in 1989 for the Trinity River, and in 2001 for the West Fork Trinity River. Therefore, using the average daily discharge allowed the analysis to encompass a larger timeframe.

agency_id	site_no	datetime	discharge.mean
USGS	8057000	1925-04-01	31.0
USGS	8057000	1925-04-02	30.0
USGS	8057000	1927-04-03	32.0

3 Exploratory Data Analysis and Wrangling

```r
## Data Wrangling
trinity.raw <-
  read.csv("../EDA_Final_Project/Data/Raw_Data/USGS_Site08057000_Flow_Raw.csv")
westfork.raw <-
  read.csv("../EDA_Final_Project/Data/Raw_Data/USGS_Site08049500_Flow_Raw.csv")

class(trinity.raw)
```

```
## [1] "data.frame"
```

```r
class(westfork.raw)
```

```
## [1] "data.frame"
```

```r
colnames(trinity.raw)
```

```
## [1] "agency_cd"              "site_no"
## [3] "datetime"               "X133517_00060_00001"
## [5] "X133517_00060_00001_cd" "X133518_00060_00002"
## [7] "X133518_00060_00002_cd" "X133519_00060_00003"
## [9] "X133519_00060_00003_cd"
```

```r
colnames(westfork.raw)
```

```
## [1] "agency_cd"              "site_no"
## [3] "datetime"               "X133233_00060_00001"
## [5] "X133233_00060_00001_cd" "X133234_00060_00002"
## [7] "X133234_00060_00002_cd" "X133235_00060_00003"
## [9] "X133235_00060_00003_cd"
```

```r
colnames(trinity.raw) <- c("agency_cd", "site_no", "datetime",
                              "discharge.max", "discharge.max.approval",
                              "discharge.min", "discharge.min.approval",
                              "discharge.mean", "discharge.mean.approval")
colnames(westfork.raw) <- c("agency_cd", "site_no", "datetime",
                              "discharge.max", "discharge.max.approval",
                              "discharge.min", "discharge.min.approval",
                              "discharge.mean", "discharge.mean.approval")

head(trinity.raw)
```

```
##   agency_cd site_no datetime discharge.max discharge.max.approval
## 1      USGS 8057000   4/1/25            NA
## 2      USGS 8057000   4/2/25            NA
## 3      USGS 8057000   4/3/25            NA
```

```
## 4      USGS 8057000   4/4/25            NA
## 5      USGS 8057000   4/5/25            NA
## 6      USGS 8057000   4/6/25            NA
##   discharge.min discharge.min.approval discharge.mean
## 1            NA                                    31
## 2            NA                                    30
## 3            NA                                    32
## 4            NA                                    30
## 5            NA                                    27
## 6            NA                                    26
##   discharge.mean.approval
## 1                       A
## 2                       A
## 3                       A
## 4                       A
## 5                       A
## 6                       A
```

```r
head(westfork.raw)
```

```
##   agency_cd site_no datetime discharge.max discharge.max.approval
## 1      USGS 8049500   4/1/25            NA
## 2      USGS 8049500   4/2/25            NA
## 3      USGS 8049500   4/3/25            NA
## 4      USGS 8049500   4/4/25            NA
## 5      USGS 8049500   4/5/25            NA
## 6      USGS 8049500   4/6/25            NA
##   discharge.min discharge.min.approval discharge.mean
## 1            NA                                    32
## 2            NA                                    27
## 3            NA                                    24
## 4            NA                                    27
## 5            NA                                    22
## 6            NA                                    28
##   discharge.mean.approval
## 1                       A
## 2                       A
## 3                       A
## 4                       A
## 5                       A
## 6                       A
```

```r
class(trinity.raw$datetime)
```

```
## [1] "factor"
```

```r
class(westfork.raw$datetime)
```

```
## [1] "factor"
```

```r
# Convert "datetime" class to date
trinity.raw$datetime <- as.Date(trinity.raw$datetime,
                                                  format = "%m/%d/%y")
westfork.raw$datetime <- as.Date(westfork.raw$datetime,
                                                  format = "%m/%d/%y")

# Fix years before 1925
trinity.raw$datetime <- format(trinity.raw$datetime, "%y%m%d")
westfork.raw$datetime <- format(westfork.raw$datetime, "%y%m%d")

create.early.dates <- (function(d) {
       paste0(ifelse(d > 190404,"19","20"),d)
       })

trinity.raw$datetime <- create.early.dates(trinity.raw$datetime)
westfork.raw$datetime <- create.early.dates(westfork.raw$datetime)

# Reformat dates
trinity.raw$datetime <- as.Date(trinity.raw$datetime,
                                                  format = "%Y%m%d")
westfork.raw$datetime <- as.Date(westfork.raw$datetime,
                                                  format = "%Y%m%d")

# Convert "site_no" class to character
trinity.raw$site_no <- as.character(trinity.raw$site_no)
westfork.raw$site_no <- as.character(westfork.raw$site_no)

# Separate "datetime" column
trinity.date <-
  separate(trinity.raw, datetime, c("Year", "Month", "Day"))
westfork.date <-
  separate(westfork.raw, datetime, c("Year", "Month", "Day"))

# Thin data to contain only mean discharge
trinity.thinned <- select(trinity.date, site_no, Year, discharge.mean)
westfork.thinned <- select(westfork.date, site_no, Year, discharge.mean)

# Remove "discharge.mean" NAs
trinity.flow <- na.omit(trinity.thinned)
westfork.flow <- na.omit(westfork.thinned)
```

```r
# Merge datasets
all.flow.data <- merge(trinity.flow, westfork.flow, all = TRUE)
```

The above R chunk contains code wrangles and tidies the datasets. After the datasets are imported, the dates are not formatted correctly and need to be converted to the proper class, as well as reformated to be usable for the analysis. The datsets are then refined to be in a useful format for the analysis.

```r
# Determine extreme event thresholds
quantile(trinity.flow$discharge.mean)
```

```
##       0%      25%      50%      75%     100%
##      8.2    250.0    479.0   1580.0 103000.0
```

```r
quantile(westfork.flow$discharge.mean)
```

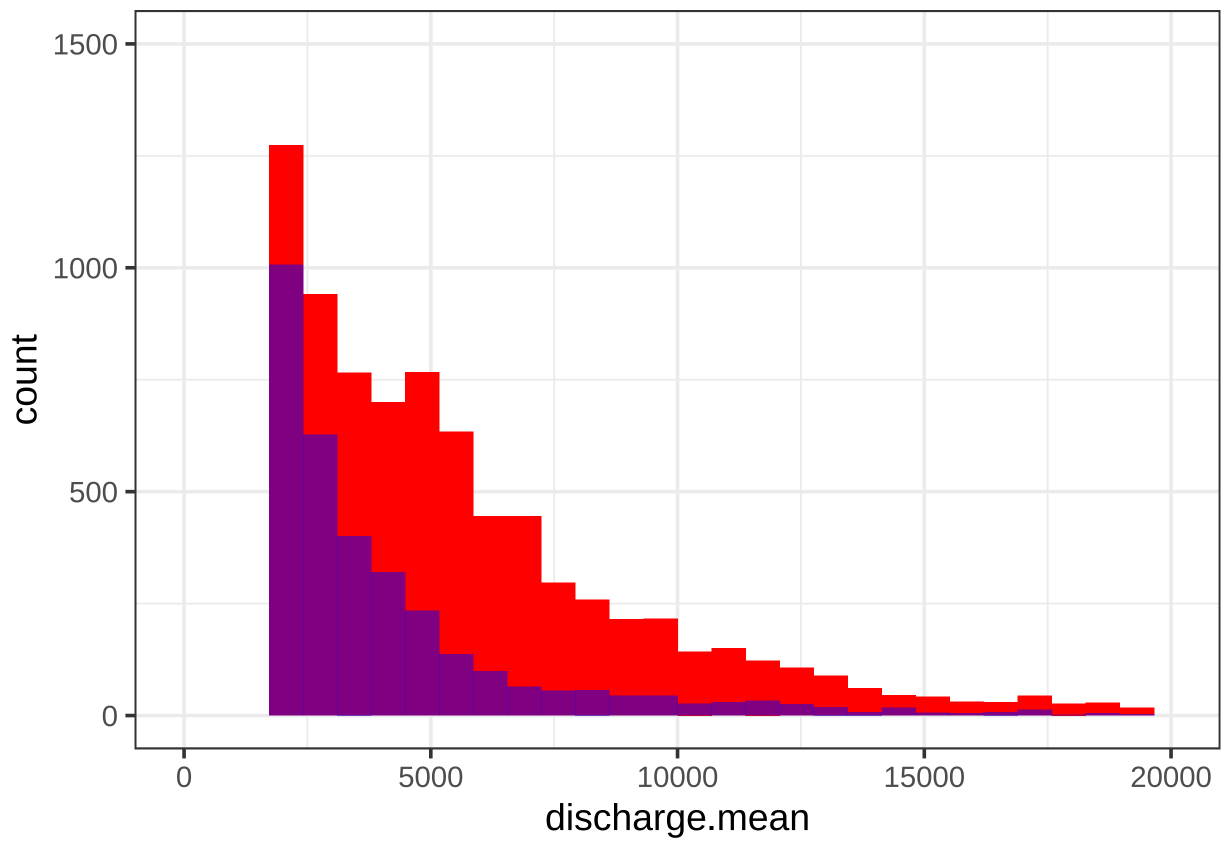
```
##       0%      25%      50%      75%     100%
##     4.50   115.00   206.00   512.25 48900.00
```

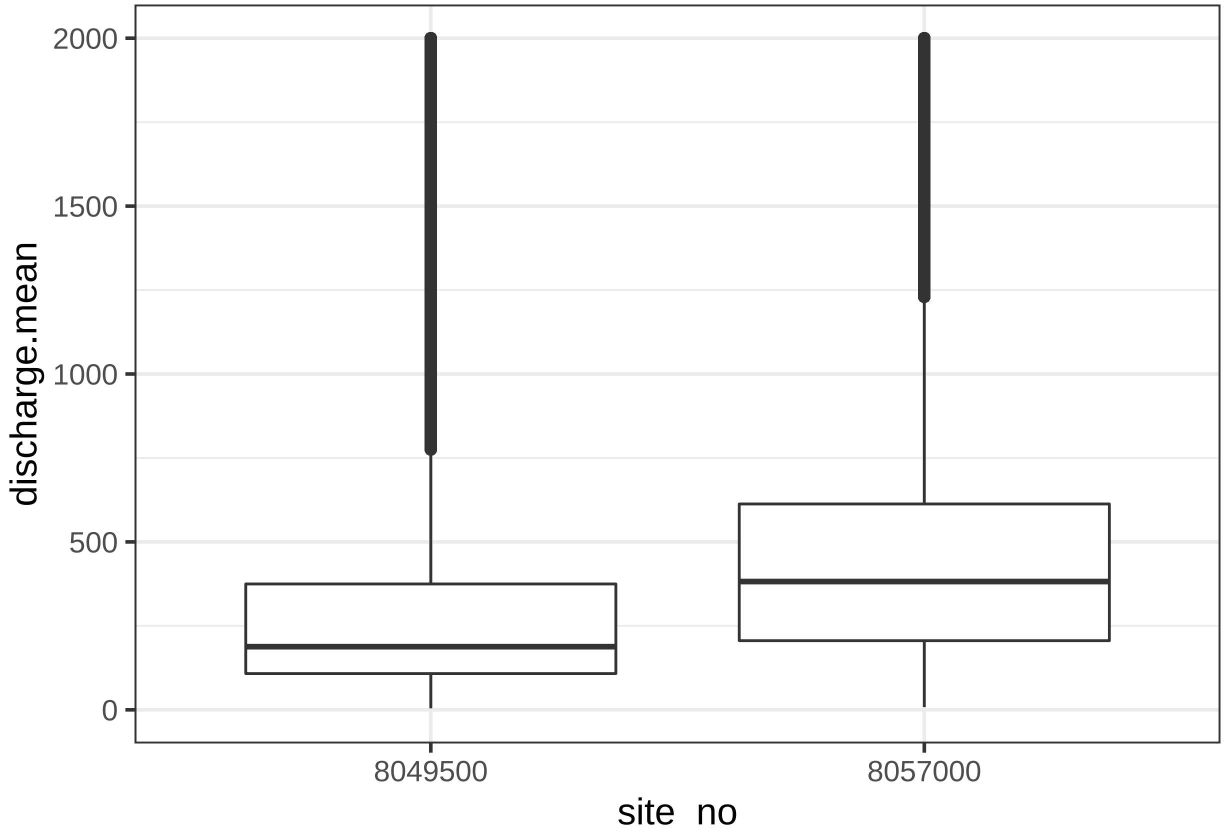
```r
# Create flood/drought data frames
trinity.flood <- filter(trinity.flow, discharge.mean >= 1580)
trinity.drought <- filter(trinity.flow, discharge.mean <= 250)
westfork.flood <- filter(westfork.flow, discharge.mean >= 512.25)
westfork.drought <- filter(westfork.flow, discharge.mean <= 115)

# Create count column of frequency of extreme events by year
trinity.flood.freq <- count(trinity.flood, Year)
trinity.drought.freq <- count(trinity.drought, Year)
westfork.flood.freq <- count(westfork.flood, Year)
westfork.drought.freq <- count(westfork.drought, Year)

# Append "site_no" column to frequency dataframes
trinity.flood.freq$site_no <- c(08057000)
trinity.drought.freq$site_no <- c(08057000)
westfork.flood.freq$site_no <- c(08049500)
westfork.drought.freq$site_no <- c(08049500)

# Rearrange columns
trinity.flood.freq <-
  select(trinity.flood.freq, site_no, Year, n)
trinity.drought.freq <-
  select(trinity.drought.freq, site_no, Year, n)
westfork.flood.freq <-
  select(westfork.flood.freq, site_no, Year, n)
westfork.drought.freq <-
  select(westfork.drought.freq, site_no, Year, n)
```

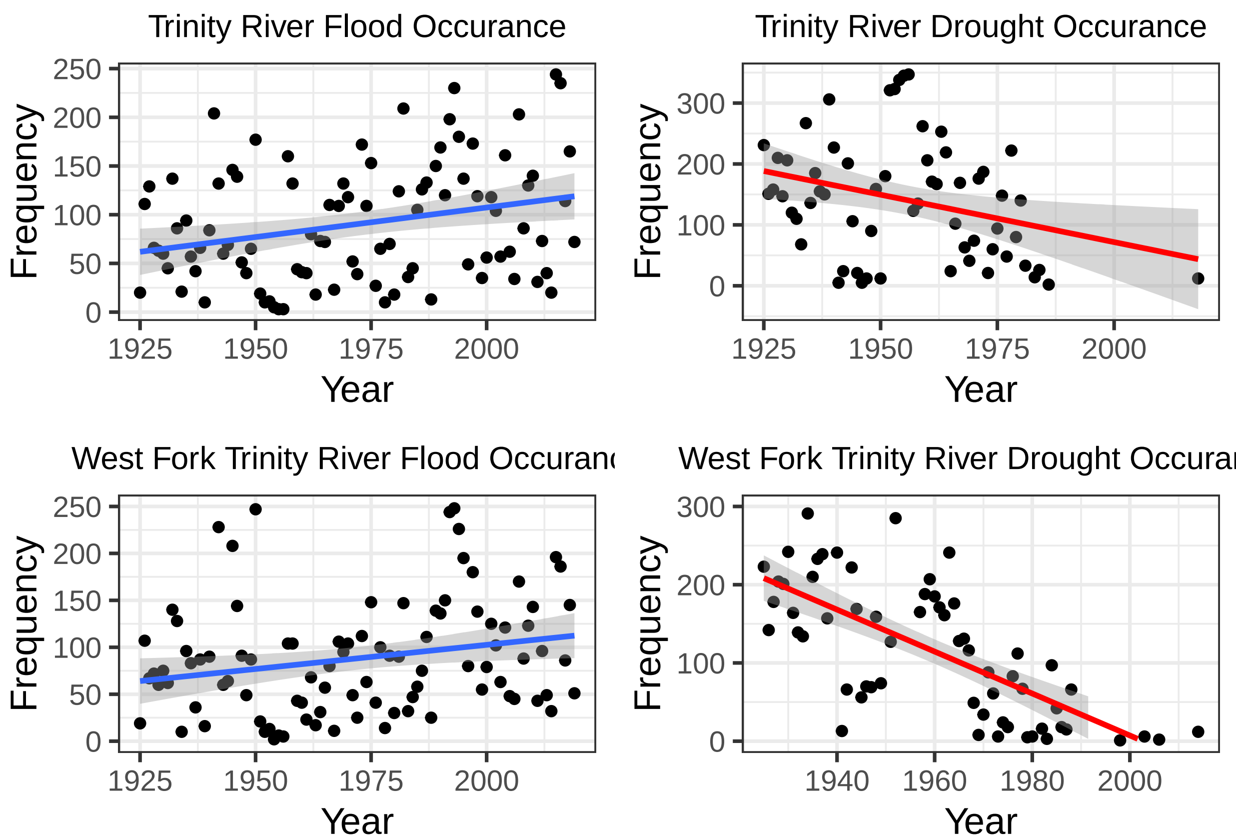
```r
# Make "Year" class numeric
trinity.flood.freq$Year <- as.numeric(trinity.flood.freq$Year)
trinity.drought.freq$Year <- as.numeric(trinity.drought.freq$Year)
westfork.flood.freq$Year <- as.numeric(westfork.flood.freq$Year)
westfork.drought.freq$Year <- as.numeric(westfork.drought.freq$Year)
```

New dataframes of flood/drought occurance by year are created for the analysis. The top and bottom quartiles are determined for each gauge, and the dataframe is further refined in order to create the most simple dataset possible for the analysis.

```
## `stat_bin()` using `bins = 30`. Pick better value with `binwidth`.
## `stat_bin()` using `bins = 30`. Pick better value with `binwidth`.
```




2000
1500
1000
500
0
discharge.mean
8049500
8057000
site_no



The exploratory graphs help orient the data, and it confirms assumptions about the data, such as the tributary river having a similar trend in discharge over time, but at a less intense magnitude. It also reveals some unexpected trends that will be tested in the statistical analysis. In particular, it reveals the trend of decreasing "drought" days and slightly increasing "flood" days.

4 Analysis

```r
# Run Man-Kendall test to test for monotonic relationship
mk.test(trinity.flood.freq$n)
```

```
##
##  Mann-Kendall trend test
##
## data:  trinity.flood.freq$n
## z = 2.2091, n = 95, p-value = 0.02717
## alternative hypothesis: true S is not equal to 0
## sample estimates:
##            S         varS          tau
## 6.880000e+02 9.671667e+04 1.544858e-01
```

```r
  # p < 0.05, trend present (z = +2.21)
mk.test(trinity.drought.freq$n)
```

```
##
##  Mann-Kendall trend test
##
## data:  trinity.drought.freq$n
## z = -2.3526, n = 61, p-value = 0.01864
## alternative hypothesis: true S is not equal to 0
## sample estimates:
##             S          varS           tau
##  -379.0000000 25815.6666667    -0.2075011
```

```r
  # p < 0.05, trend present (z = -2.35)
mk.test(westfork.flood.freq$n)
```

```
##
##  Mann-Kendall trend test
##
## data:  westfork.flood.freq$n
## z = 2.1318, n = 95, p-value = 0.03302
## alternative hypothesis: true S is not equal to 0
## sample estimates:
##            S         varS          tau
## 6.640000e+02 9.672133e+04 1.490296e-01
```

```r
  # p < 0.05, trend present (z = +2.13)
mk.test(westfork.drought.freq$n)
```

```
##
##  Mann-Kendall trend test
```

```
##
## data:  westfork.drought.freq$n
## z = -5.745, n = 66, p-value = 9.193e-09
## alternative hypothesis: true S is not equal to 0
## sample estimates:
##             S          varS           tau
## -1039.0000000 32645.0000000    -0.4850612
```

```
  # p < 0.05, trend present (z = -5.75)

# Run Pettitt's test
pettitt.test(trinity.flood.freq$n)
```

```
##
##  Pettitt's test for single change-point detection
##
## data:  trinity.flood.freq$n
## U* = 804, p-value = 0.02274
## alternative hypothesis: two.sided
## sample estimates:
## probable change point at time K
##                              56
```

```
  # Change point @ row 56 (1980)
pettitt.test(trinity.drought.freq$n)
```

```
##
##  Pettitt's test for single change-point detection
##
## data:  trinity.drought.freq$n
## U* = 424, p-value = 0.01864
## alternative hypothesis: two.sided
## sample estimates:
## probable change point at time K
##                              40
```

```
  # Change point @ row 40 (1964)
pettitt.test(westfork.flood.freq$n)
```

```
##
##  Pettitt's test for single change-point detection
##
## data:  westfork.flood.freq$n
## U* = 879, p-value = 0.00949
## alternative hypothesis: two.sided
## sample estimates:
## probable change point at time K
```

```
##                                 64
```

```r
  # Change point @ row 64 (1988)
pettitt.test(westfork.drought.freq$n)
```

```
##
##  Pettitt's test for single change-point detection
##
## data:  westfork.drought.freq$n
## U* = 934, p-value = 3.253e-08
## alternative hypothesis: two.sided
## sample estimates:
## probable change point at time K                              <NA>
##                                 39                              41
```

```r
  # Change point @ row 39 (1964)

# Run separate Man-Kendall for each change point
mk.test(trinity.flood.freq$n[1:55])
```

```
##
##  Mann-Kendall trend test
##
## data:  trinity.flood.freq$n[1:55]
## z = -0.29048, n = 55, p-value = 0.7714
## alternative hypothesis: true S is not equal to 0
## sample estimates:
##             S          varS           tau
## -4.100000e+01  1.896167e+04 -2.772166e-02
```

```r
  # p > 0.05, no trend
mk.test(trinity.flood.freq$n[56:95])
```

```
##
##  Mann-Kendall trend test
##
## data:  trinity.flood.freq$n[56:95]
## z = 0.011651, n = 40, p-value = 0.9907
## alternative hypothesis: true S is not equal to 0
## sample estimates:
##            S         varS          tau
## 2.000000e+00 7.366667e+03 2.564103e-03
```

```r
  # p > 0.05, no trend
mk.test(trinity.drought.freq$n[1:39])
```

```
##
##  Mann-Kendall trend test
```

```
##
## data:  trinity.drought.freq$n[1:39]
## z = 1.0043, n = 39, p-value = 0.3153
## alternative hypothesis: true S is not equal to 0
## sample estimates:
##           S        varS         tau
##   84.0000000 6830.6666667    0.1135905
```

```r
  # p > 0.05, no trend
mk.test(trinity.drought.freq$n[40:61])
```

```
##
##  Mann-Kendall trend test
##
## data:  trinity.drought.freq$n[40:61]
## z = -2.0866, n = 22, p-value = 0.03692
## alternative hypothesis: true S is not equal to 0
## sample estimates:
##           S        varS         tau
##  -75.0000000 1257.6666667   -0.3246753
```

```r
  # p < 0.05, trend present (z = -2.09)
mk.test(westfork.flood.freq$n[1:63])
```

```
##
##  Mann-Kendall trend test
##
## data:  westfork.flood.freq$n[1:63]
## z = -0.26695, n = 63, p-value = 0.7895
## alternative hypothesis: true S is not equal to 0
## sample estimates:
##             S          varS           tau
## -4.600000e+01  2.841533e+04 -2.362012e-02
```

```r
  # p > 0.05, no trend
mk.test(westfork.flood.freq$n[64:95])
```

```
##
##  Mann-Kendall trend test
##
## data:  westfork.flood.freq$n[64:95]
## z = -1.6379, n = 32, p-value = 0.1015
## alternative hypothesis: true S is not equal to 0
## sample estimates:
##           S        varS         tau
## -102.0000000 3802.6666667   -0.2056452
```

```r
  # p > 0.05, no trend
mk.test(westfork.drought.freq$n[1:38])
```

```
##
##  Mann-Kendall trend test
##
## data:  westfork.drought.freq$n[1:38]
## z = 0.56578, n = 38, p-value = 0.5715
## alternative hypothesis: true S is not equal to 0
## sample estimates:
##            S         varS          tau
## 4.600000e+01 6.326000e+03 6.548044e-02
```

```r
  # p > 0.05, no trend
mk.test(westfork.drought.freq$n[39:66])
```

```
##
##  Mann-Kendall trend test
##
## data:  westfork.drought.freq$n[39:66]
## z = -3.421, n = 28, p-value = 0.0006239
## alternative hypothesis: true S is not equal to 0
## sample estimates:
##            S         varS          tau
## -174.0000000 2557.3333333   -0.4627725
```

```r
  # P < 0.05, trend present (z = -3.42)

# Test for second change point
pettitt.test(trinity.drought.freq$n[40:61])
```

```
##
##  Pettitt's test for single change-point detection
##
## data:  trinity.drought.freq$n[40:61]
## U* = 77, p-value = 0.08188
## alternative hypothesis: two.sided
## sample estimates:
## probable change point at time K
##                              17
```

```r
  # p > 0.05, no change point
pettitt.test(westfork.drought.freq$n[39:66])
```

```
##
##  Pettitt's test for single change-point detection
##
```

```
## data:  westfork.drought.freq$n[39:66]
## U* = 132, p-value = 0.02014
## alternative hypothesis: two.sided
## sample estimates:
## probable change point at time K
##                              15
```

```r
  # p < 0.05, change point @ row 44 (1969)

# Second MK test for second change point
mk.test(westfork.drought.freq$n[39:43])
```

```
##
##  Mann-Kendall trend test
##
## data:  westfork.drought.freq$n[39:43]
## z = -1.7146, n = 5, p-value = 0.08641
## alternative hypothesis: true S is not equal to 0
## sample estimates:
##        S     varS      tau
## -8.00000 16.66667 -0.80000
```

```r
mk.test(westfork.drought.freq$n[44:66])
```

```
##
##  Mann-Kendall trend test
##
## data:  westfork.drought.freq$n[44:66]
## z = -1.693, n = 23, p-value = 0.09045
## alternative hypothesis: true S is not equal to 0
## sample estimates:
##            S          varS           tau
##  -65.0000000 1429.0000000   -0.2589724
```

```r
  # p > 0.05 for both, no trends

# Run Spearman's Rho test (non-parametric)
cor.test(trinity.flood.freq$n, trinity.flood.freq$Year,
         method = "spearman", exact = FALSE)
```

```
##
##  Spearman's rank correlation rho
##
## data:  trinity.flood.freq$n and trinity.flood.freq$Year
## S = 108870, p-value = 0.02017
## alternative hypothesis: true rho is not equal to 0
## sample estimates:
```

```
##       rho
## 0.2380606
```

```r
 # p < 0.05 (0.02), rho = 0.238, weak correlation
cor.test(trinity.drought.freq$n, trinity.drought.freq$Year,
         method = "spearman", exact = FALSE)
```

```
##
##  Spearman's rank correlation rho
##
## data:  trinity.drought.freq$n and trinity.drought.freq$Year
## S = 48975, p-value = 0.02102
## alternative hypothesis: true rho is not equal to 0
## sample estimates:
##        rho
## -0.2949545
```

```r
 # p < 0.05 (0.02), rho = -0.295, weak correlation
cor.test(westfork.flood.freq$n, westfork.flood.freq$Year,
         method = "spearman", exact = FALSE)
```

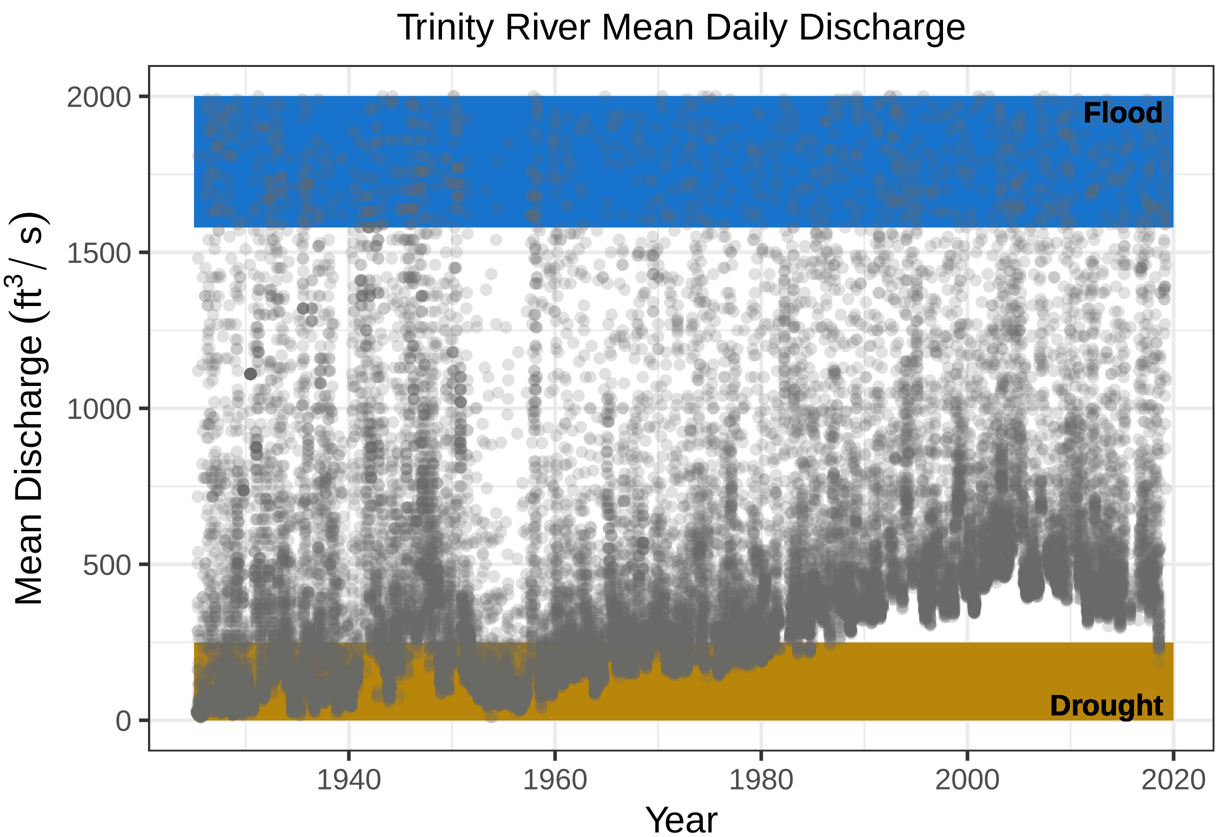
```
##
##  Spearman's rank correlation rho
##
## data:  westfork.flood.freq$n and westfork.flood.freq$Year
## S = 112320, p-value = 0.0374
## alternative hypothesis: true rho is not equal to 0
## sample estimates:
##       rho
## 0.2138945
```

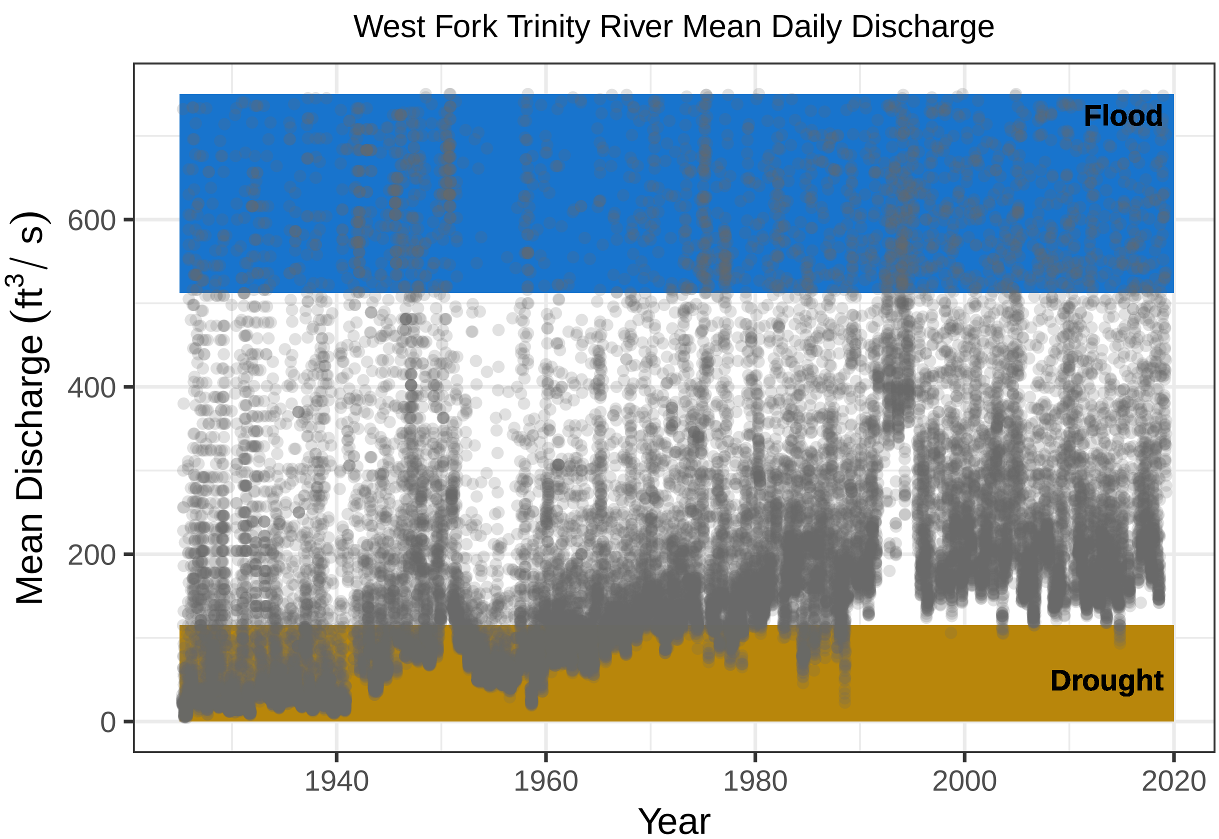
```r
  # p < 0.05 (0.04), rho = 0.214, weak correlation
cor.test(westfork.drought.freq$n, westfork.drought.freq$Year,
         method = "spearman", exact = FALSE)
```

```
##
##  Spearman's rank correlation rho
##
## data:  westfork.drought.freq$n and westfork.drought.freq$Year
## S = 80506, p-value = 3.255e-10
## alternative hypothesis: true rho is not equal to 0
## sample estimates:
##        rho
## -0.6805424
```

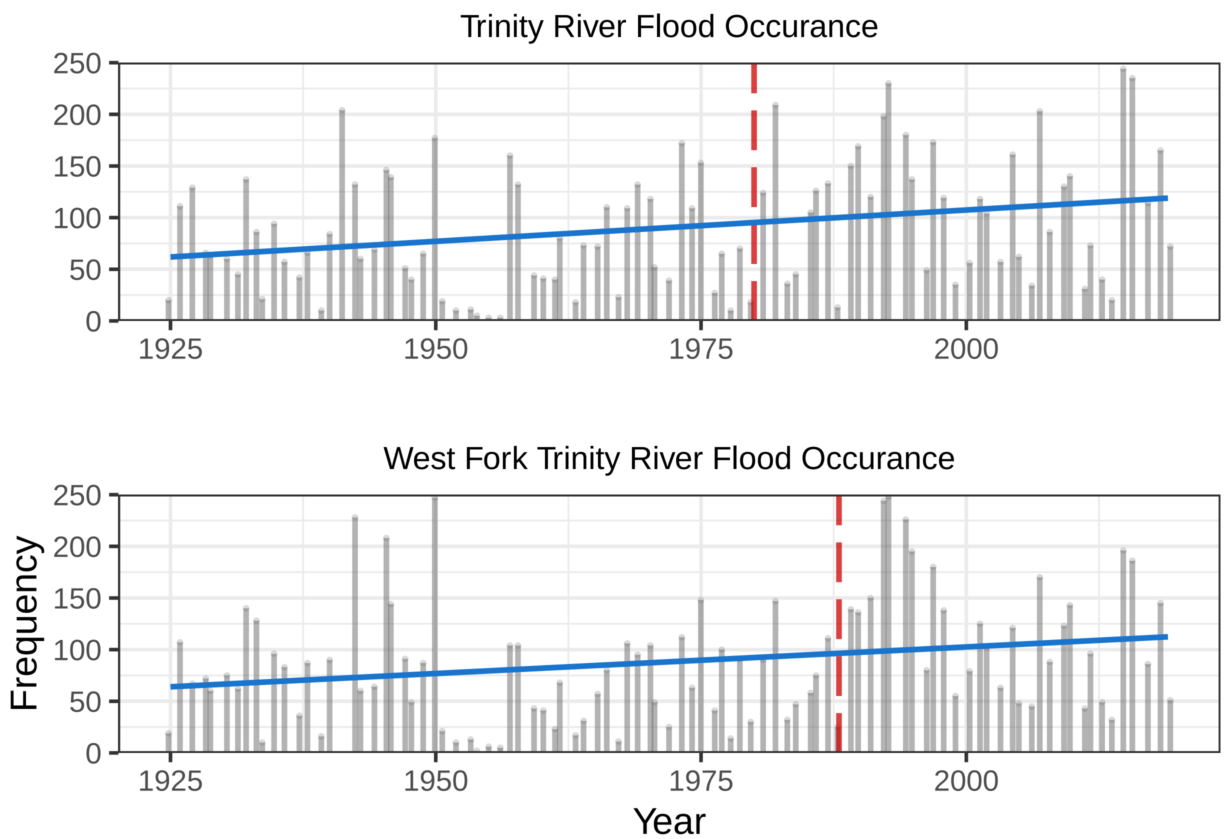
```r
  # p < 0.05 (3.2^-10), rho = -0.681, strong correlation
```

First, a Mann_Kendall test was run to determine if the trends in discharge are monotonic. A Pettitt's test was then run on each dataset to see if there were any change points present, which could be indicative of increased river control and/or dramatic climatic changes. Because all four trends are monotonic, albeit weakly monotonic, a Spearman Rho's test was run to determine the strength of the correlation between the extreme hydrologic events and time.

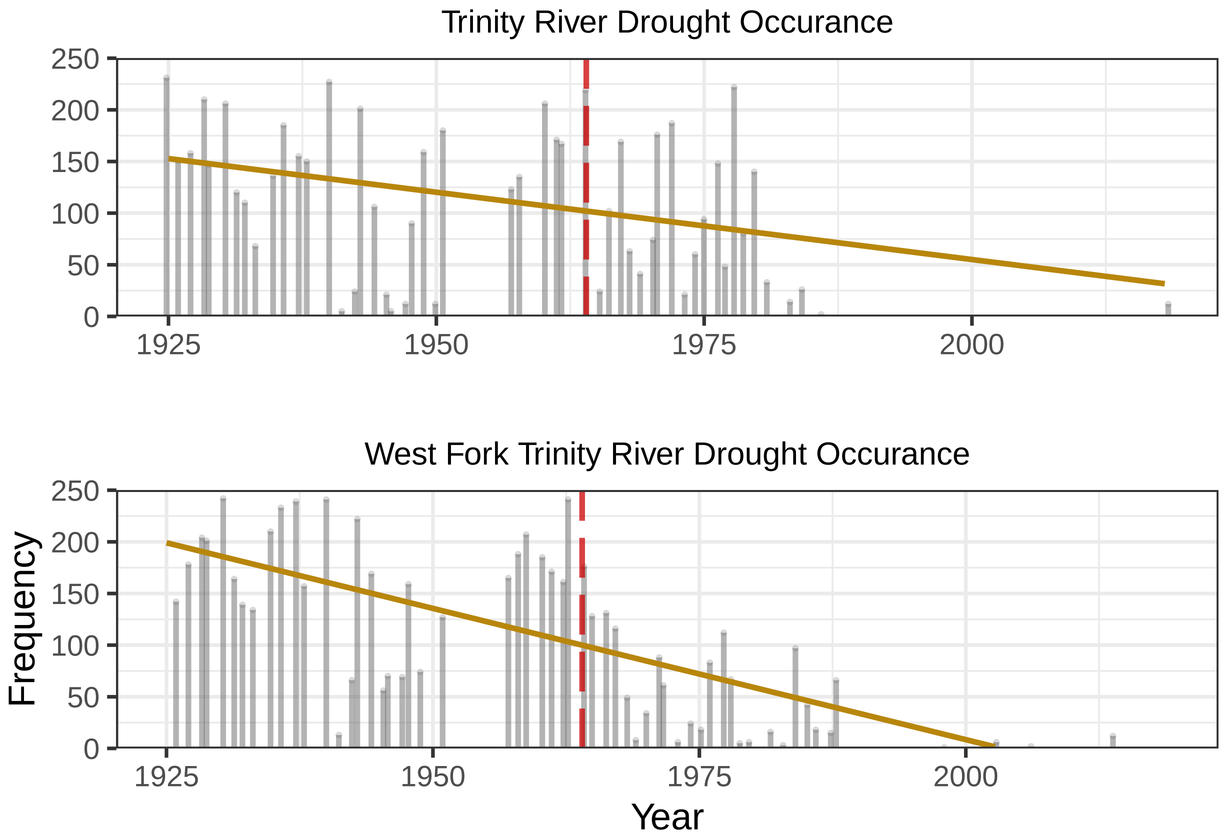




Figures 1 and 2 visually dipict how often the discharges of both rivers fall into the top and bottom quintiles, characterized as "Drought" for the bottom 25% and "Flood" for the top 25%. Both of the above figures were rescaled to better visualize the thresholds for flood and drought discharges for each river. 7,615 values above 2,000 were removed from Figure 2, and 6,645 values above 750 were removed from Figure 3.



Trinity River Flood Occurance
West Fork Trinity River Flood Occurance
250
200
150
100
50
0
1925
1950
1975
2000
Frequency
Year



Figures 3 and 4 drive home the main takeaway from the analysis- "drought" discharges have decreased, and "flood" discharges have slightly increased over time in the Dallas area. Red dashed lines indicate change points, which provide additional information about when infrastructural or climatic changes began to really affect the surface water hydrology of the Dallas area.

5 Summary and Conclusions

The discharge of rivers in the Dallas area show that the region is getting wetter over time. In both the Trinity River and the West Fork Trinity River, “droughts” have become less frequent, while “floods” have become more frequent. However, even though the Mann-Kendall tests revealed that these trends are indeed monotonic, none of the trends are particularlly strong. The strongest correlation between extreme discharge frequency and time were “drought” discharge conditions in the West Fork Trinity River, which can be characterized as a moderately strong negative trend based on the results of the Man-Kendall test ($z = -5.75$). The other Man-Kendall tests produced absolute z values no greater than 2.35, indiicating that while the trends are monotonic, they are relatively weak trends.

Pettitt’s test was used to determine if any change points exist within the data. A change point is where the magnitude or direction of the trend changes in the data. The analysis found that in both rivers, a change point in “flood” discharges occurred in the 1980s, and the change point in the “drought” discharges occurred in 1964. The only significant trends before or after the change point were found in the “drought” datasets for both rivers, which experienced negative trends after 1964. Meaning, both the Trinity River and the West Fork Trinity River experienced less frequent “drought” discharges after 1964.

The results of this analysis indicate that the Trinity River and it’s tributary, the West Fork Trinity River, are experiencing similar trends in the frequency of extreme hydrologic events. However, the trends in the larger Trinity River are weaker overall than the West Fork Trinity River, particularlly in “drought” discharge frequency. Maintaing minimum flows in the Trinity River is much more important than controling flows in the West Fork Trinity River, considering the Trinity River is the largest river in the area and is used as a source of drinking water. The analysis confirms that recent extreme events in the Dallas area are indicative of an emerging trend. It also confirms that extreme discharge trends in a larger, more regulated river like the Trinity River are less intense than a lower order tributary.