

Spike Challenge

Yessenia Valencia Perez

17/10/2019

Analisis del dataset

```
codigo<-table(Data$codigo_estacion)
nombre<-table(Data$nombre)
```

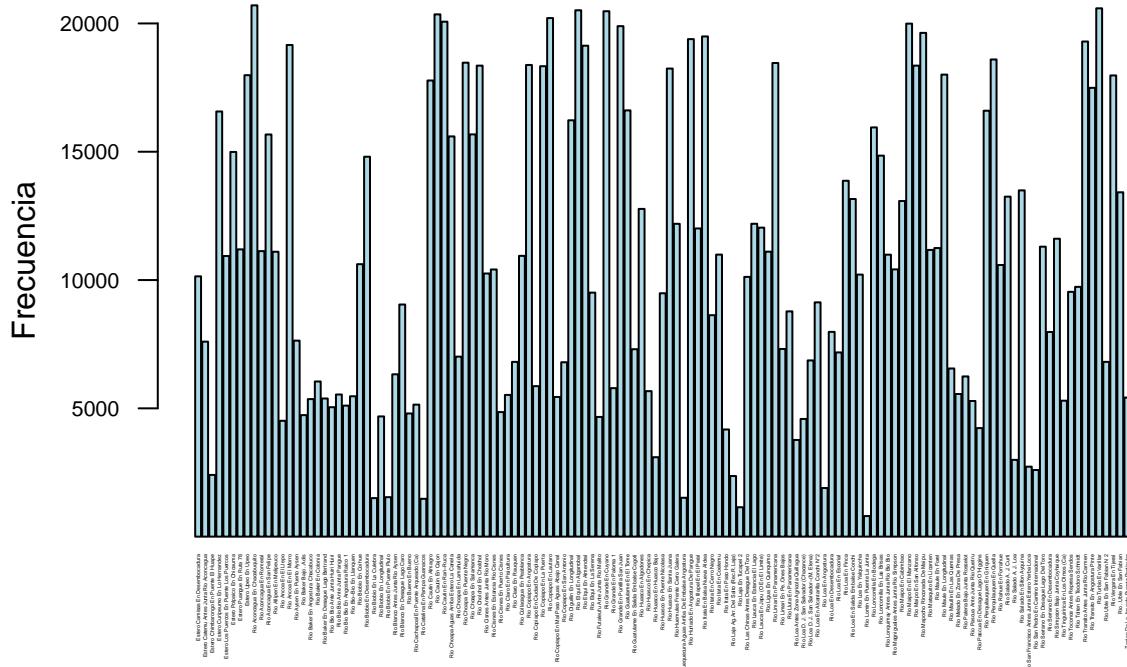
La medicion del caudal esta compuesta por 133 distintos codigos de la estacion.

```
length(codigo)
```

```
## [1] 133
```

La mayoria de las estaciones se encuentran en el Rio Aconcagua En Chacabuquito y en el Rio Lomin En Puente La Junta es donde menos estaciones hay.

```
barplot(nombre,ylab="Frecuencia",xlab="",col="lightblue",
       las=2,cex.axis = 0.7,cex.names = 0.2,ylim = c(802,20706))
```

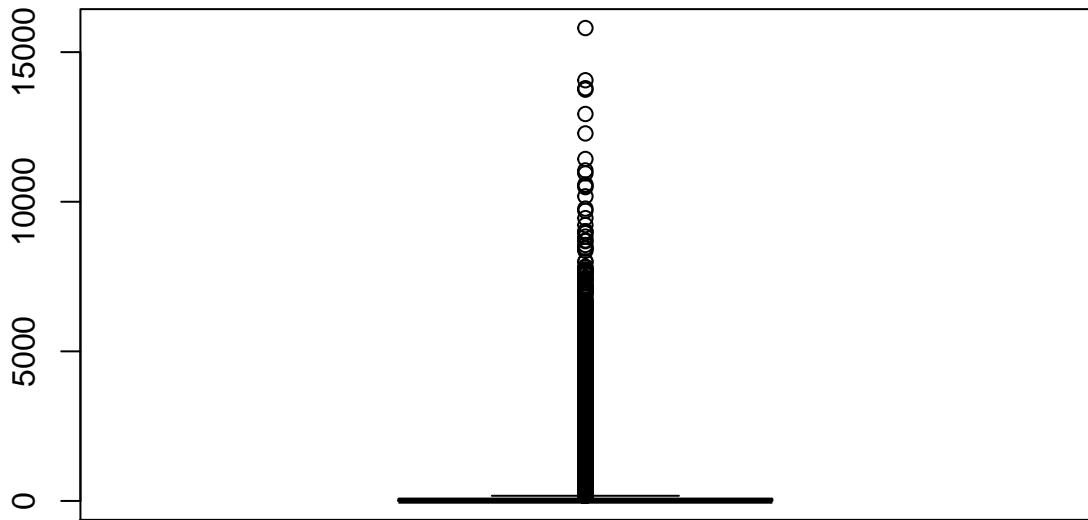


En cuanto a las mediciones del caudal es posible notar que estas varian entre 0 y 15805, donde su media y mediana difieren, lo que nos indica que los datos se distribuyen de forma asimetrica positiva. Ademas, existe una gran cantidad de datos extremos lo que indicaria ausencia de gausianidad.

```
summary(caudal)
```

```
##      Min.    1st Qu.     Median      Mean    3rd Qu.      Max.
##      0.00     1.25     8.95   95.52   69.70  15805.00
```

```
boxplot(caudal,xlab="Medicion del caudal")
```



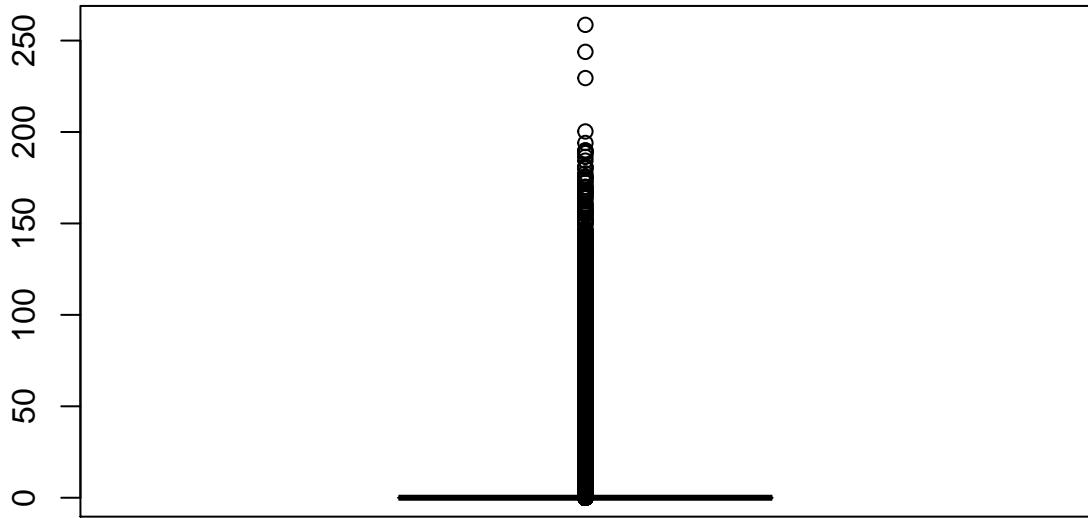
Medicion del caudal

Las precipitaciones promedio en la cuenca varian entre 0 y 258.6, donde se observan 27767 valores missing y una gran cantidad de valores extremos.

```
summary(precipitacion)
```

```
##      Min. 1st Qu. Median     Mean 3rd Qu.    Max.    NA's
##      0.000  0.000  0.000   1.921  0.000 258.600 27767
```

```
boxplot(precipitacion,xlab="Precipitacion promedio")
```



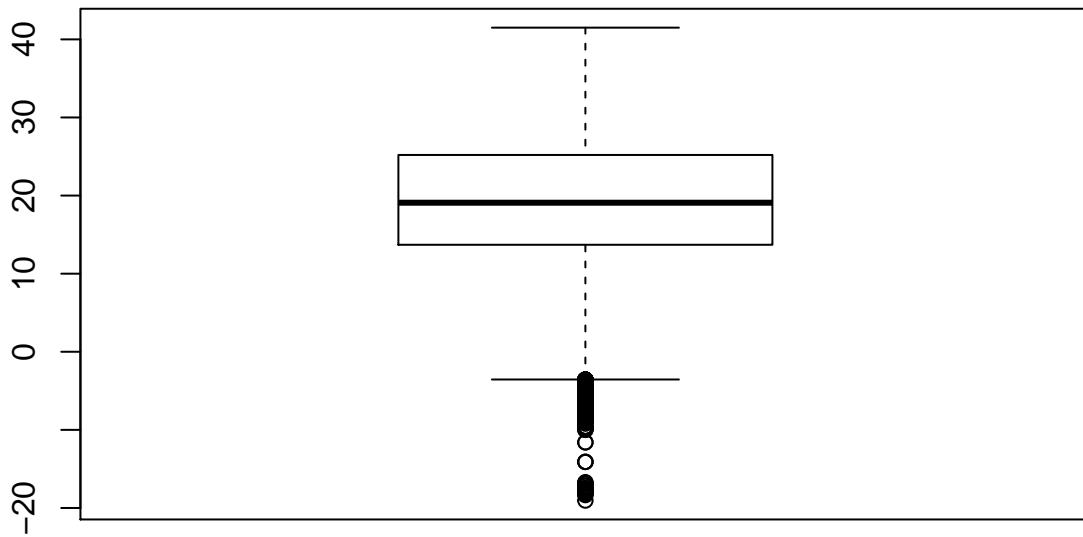
Precipitacion promedio

Las temperaturas promedio en las cuencas varian entre -19.05 y 41.5, la mediana y la media son similares y se podr??a suponer que los datos son simetricos, aunque a partir del boxplot se observa una gran cantidad de puntos extremos y ademas se determina que hay 151563 valores missing existentes en esta variable.

```
summary(temperatura)
```

```
##      Min. 1st Qu. Median     Mean 3rd Qu.     Max.    NA's
## -19.05   13.70  19.08  19.21  25.20   41.50 151563
```

```
boxplot(temperatura,xlab="Temperatura maxima")
```

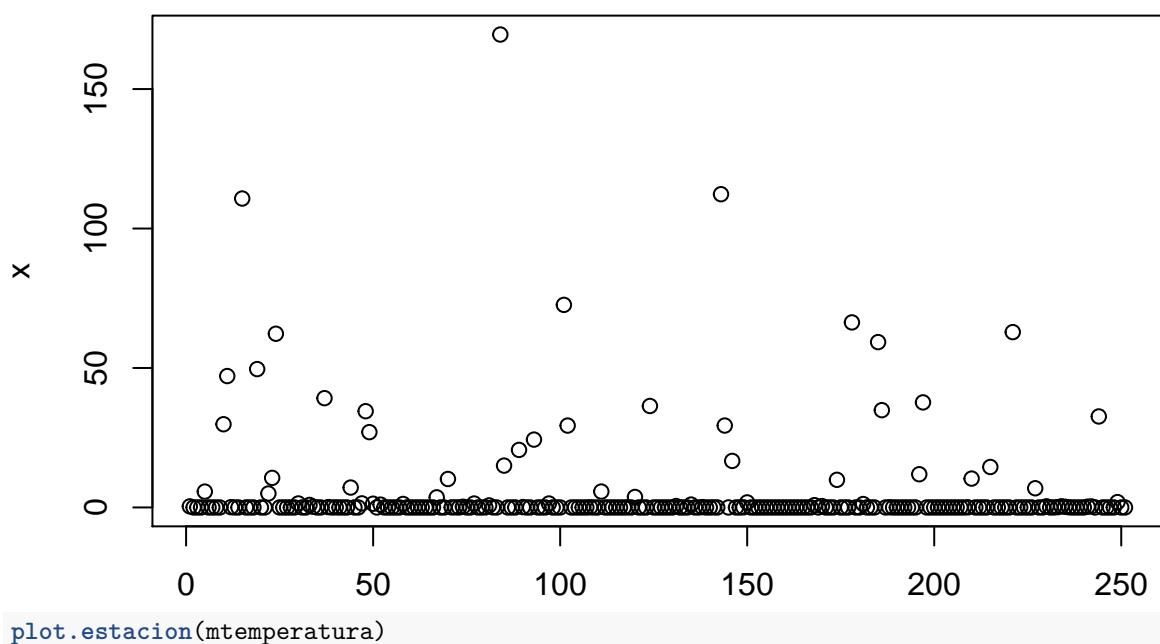
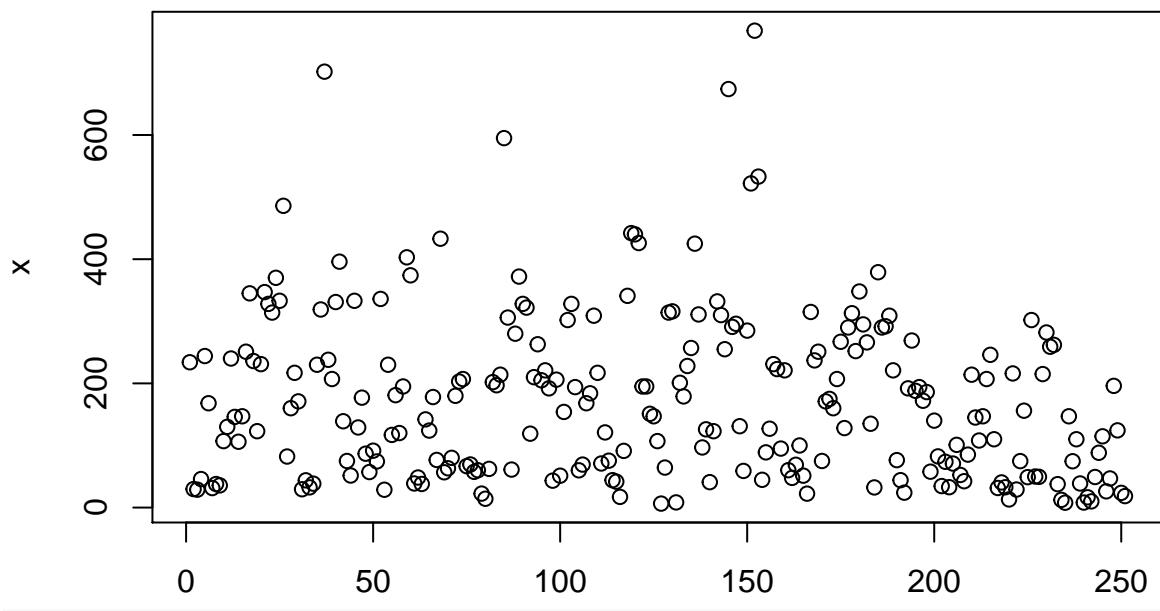


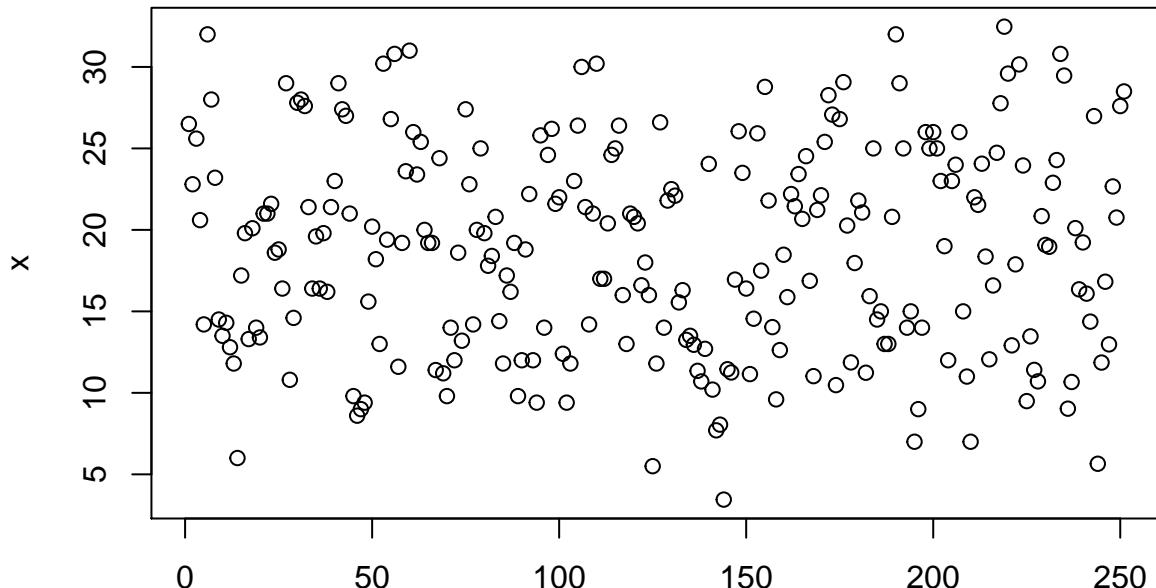
Temperatura maxima

En general las variables de interes del dataset posee una gran cantidad de valores faltantes, que probablemente no hayan sido medidos.

Plots de precipitacion, temperatura y caudal

```
plot.estacion=function(x){
  x=plot(x,xlab = "")}
plot.estacion(mcaudal)
```





Nuevas variables:

- caudal_extremo
- temp_extremo
- precip_extremo

```
variable_extrema=function(x){
ifelse(x>quantile(x,0.95),1,0)}

caudal_extremo<-variable_extrema(caudal)
temp_extremo<-variable_extrema(temperatura)
precip_extremo<-variable_extrema(precipitacion)
```

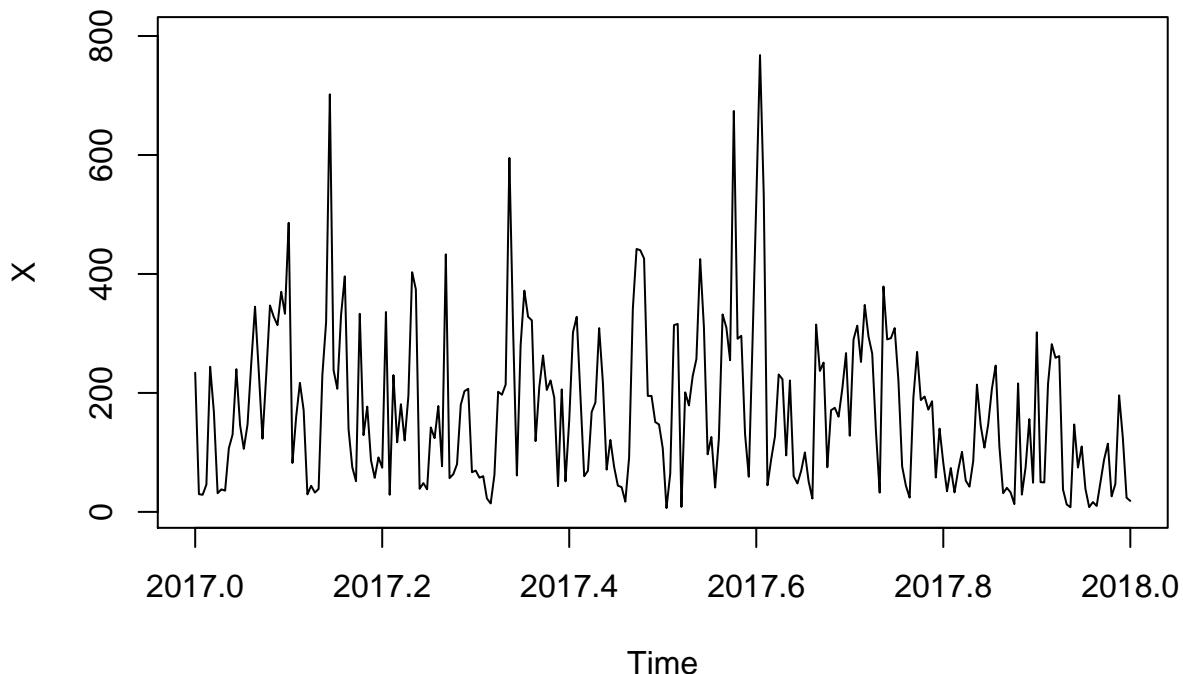
De cierto modo se cree que si es adecuada la forma de determinar un valor extremo, pero debido al cambio climatico en el ultimo tiempo es posible notar dos estaciones mas presentes, las cuales son verano e invierno, por lo que tambien se podria comparar por semestres los valores extremos.

Variable caudal_extremo

En general el comportamiento de las cuencas no son tan diferentes entre si, debido a la forma en que se ha clasificado si el caudal corresponde o no a un valor extremo, ya que el 95% en el dataset historico no fue dia extremo.

Plot del porcentaje de eseventos extremos.

```
plot(X,ylim=c(5,800))
```



En el ultimo tiempo se observa que hay una variabilidad mas constante, ya que los eventos extremo han sido mas contantes respecto a otros tiempo.

Prediccion de caudal extremo

Ya que en el clima del ultimo tiempo varia con respecto al ultimo tiempo, para realizar la prediccion solo utilizaria los datos de a lo mas 12 o 24 meses anteriores, para tomar en cuenta la variabilidad de los ultimos meses e incluirla en el modelo, tambien considero necesario utilizar la informacion del dia anterior, ya que esta tendra estricta relacion a como podria variar el caudal de dias futuros en caso de estar en presencia de una serie ciclica. Debido a que los datos trabajados son dependientes al tiempo en que sean medidos, la forma mas adecuada de generar un modelo podria ser a traves de un modelo ARMA de series de tiempo.