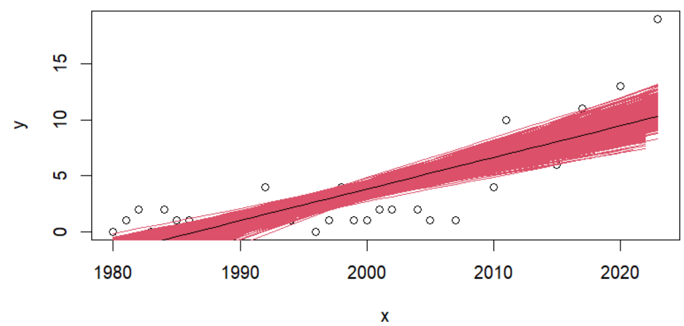
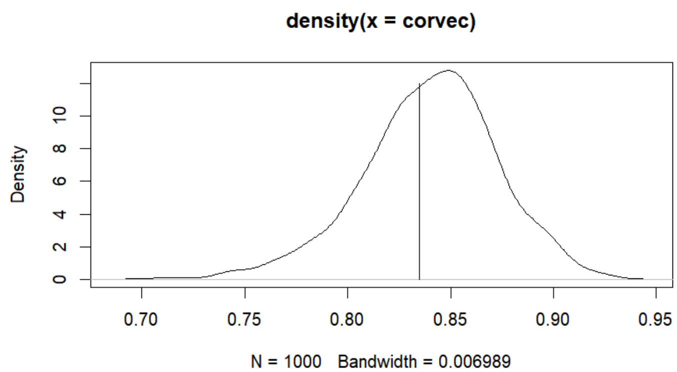
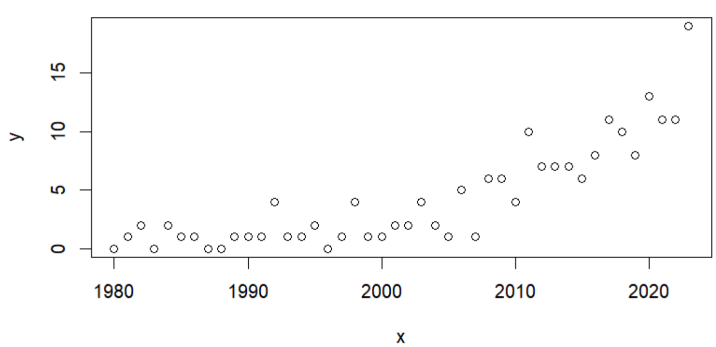
Starting regression:

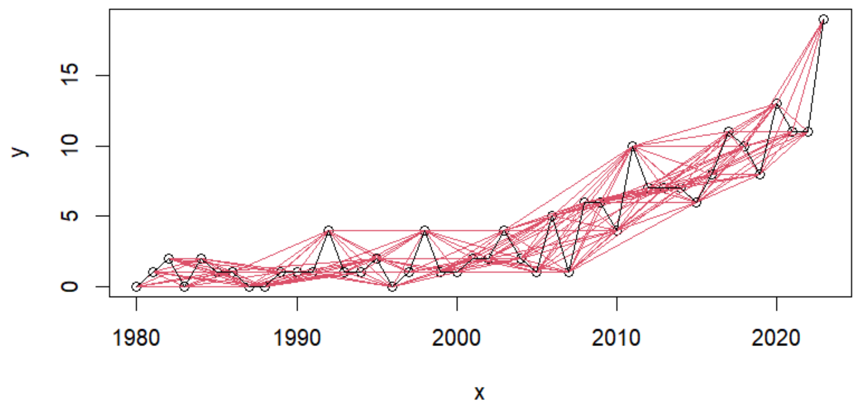
Comparing Disaster Statistics against Time:

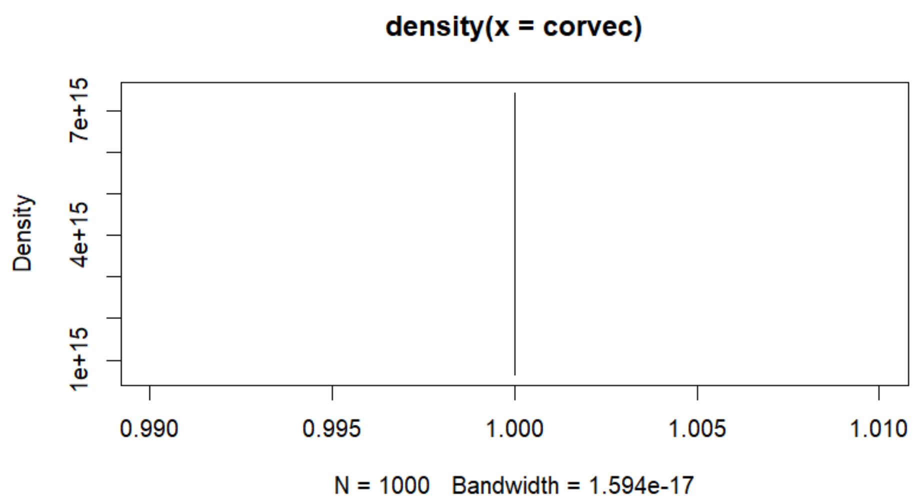
Severe Storms:





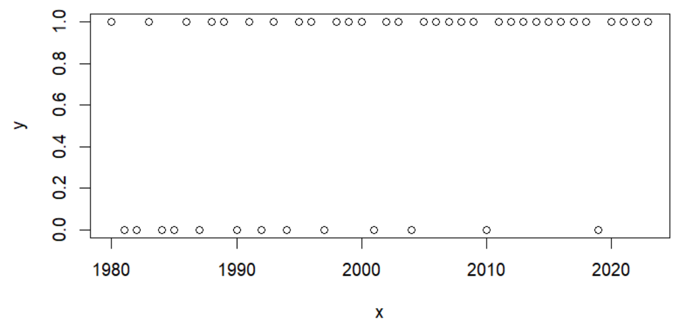


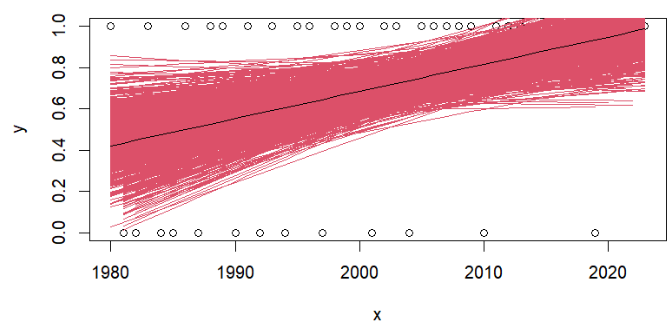


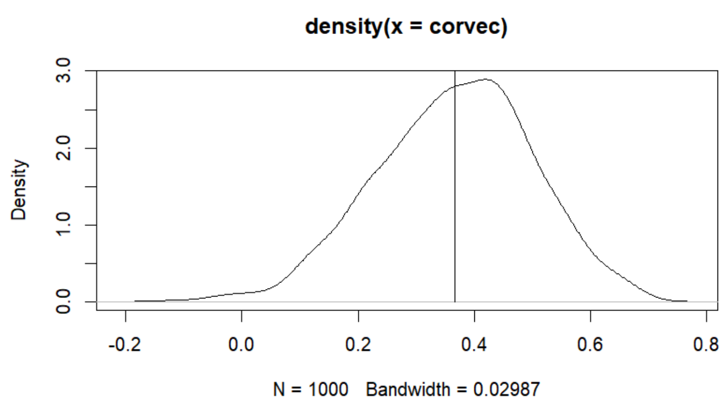


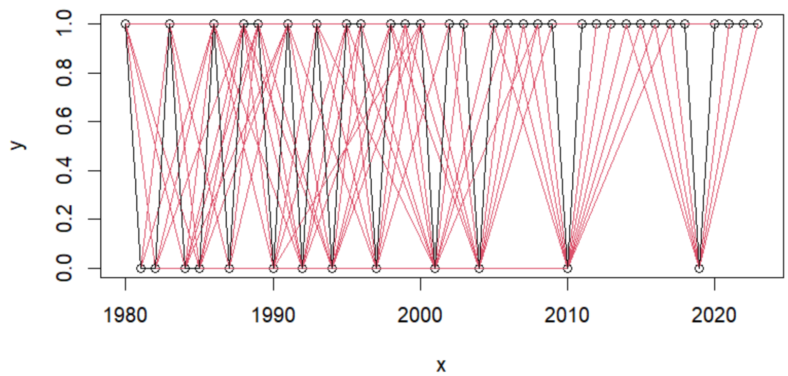
The scatter plots and the connected line graph with bootstrap replicates show that there is definitely a clear relationship between the number of severe storms and time passing. However, the smoothed density plot for the correlation shows that the correlation is centered around 0.85, which isn't ideal. Moreover, the linear regression plot with the bootstrap replicates shows that while a linear relationship could be conceived, it won't be particularly accurate and that the relationship between these variables might be non-linear.

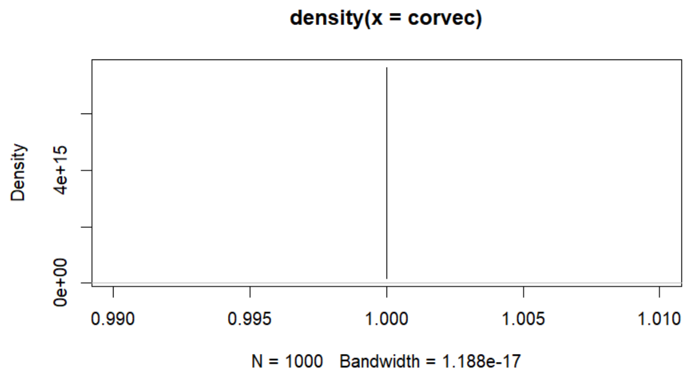
Droughts:





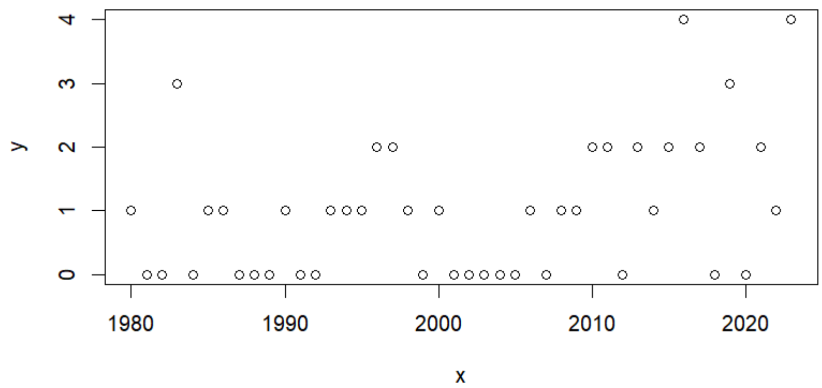


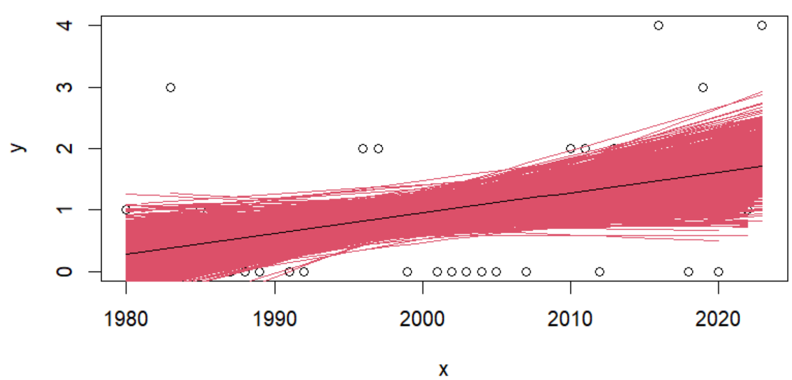


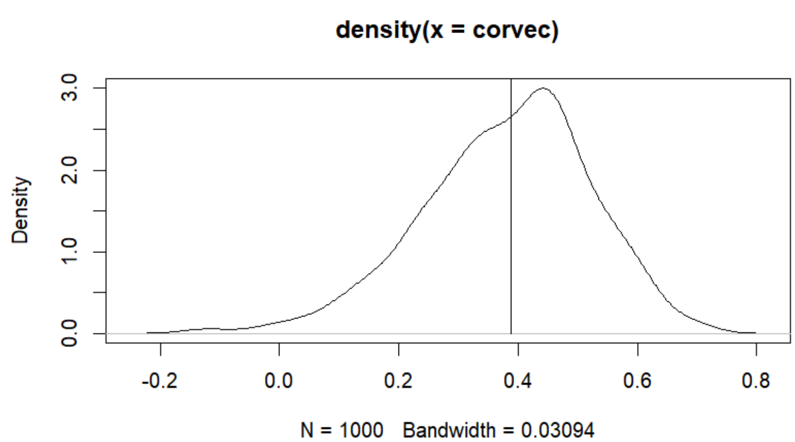


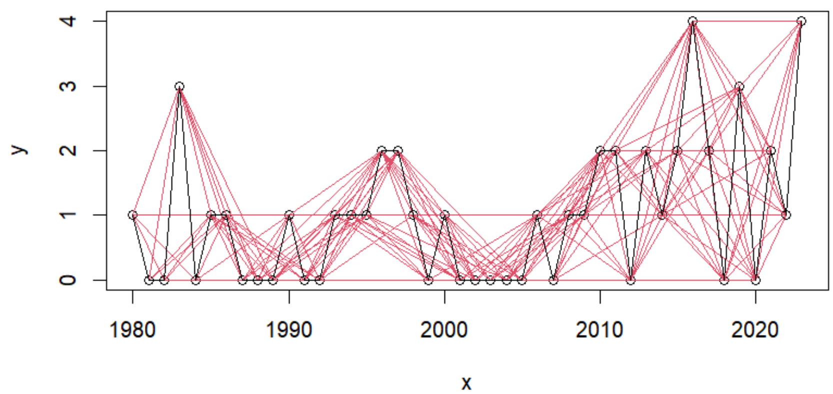
These plots reflect that the data seems to have two distinct values, 0 and 1 and that it is challenging to establish a relationship between years passed and the number of droughts. There does seem to be some relationship when we look at the line chart with the bootstrapped replicates since years without droughts seem to get more and more spaced out as the years increase. However, it is clear from the plots that any linear relationship between these variables would be non-functional, especially considering the correlation peak seems to be around 0.5, which shows a very weak linear relationship.

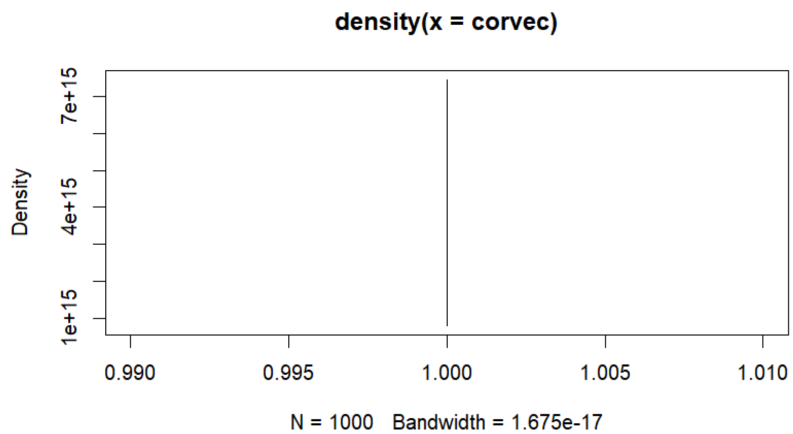
Floods:


y
0
1
2
3
4
1980
1990
2000
2010
2020
x


y
0
1
2
3
4
1980
1990
2000
2010
2020
x

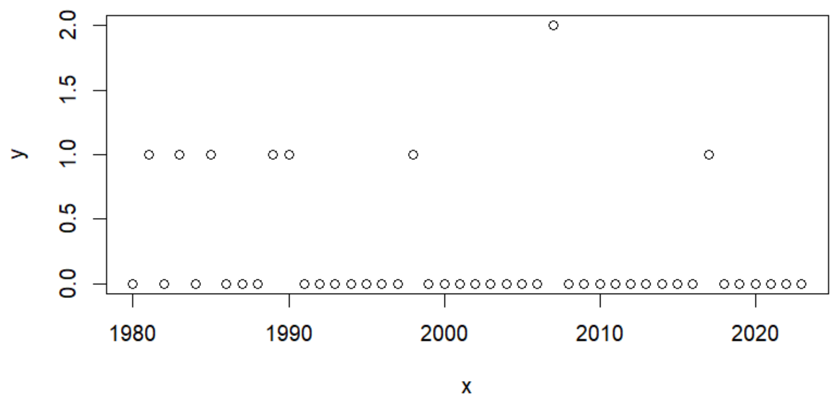

density(x = corvec)
Density
0.0
1.0
2.0
3.0
-0.2
0.0
0.2
0.4
0.6
0.8
N = 1000 Bandwidth = 0.03094



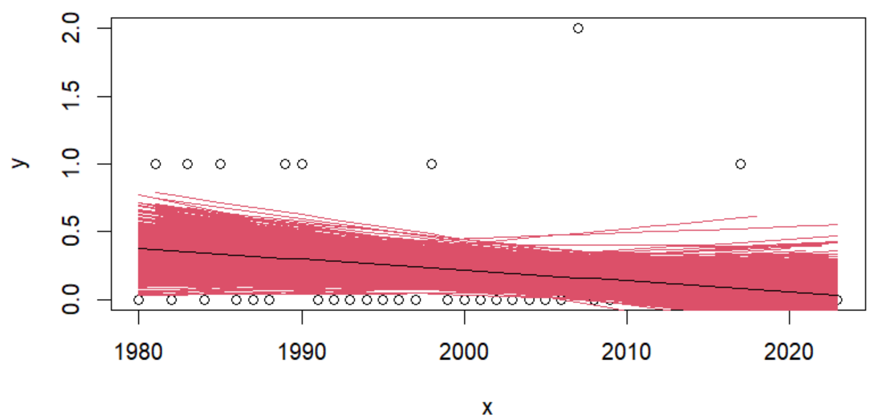


The scatter plot and the line chart show that the relationship between floods and time seems to exist but is not very strong. The relationship indicates that floods have generally increased, with one outlier between 1980 and 1990. However, the relationship very clearly does not seem linear, as is evident by the line of best-fit graph and the correlation plot, which appears to peak close to 0.5. Hence, a linear relationship would be non-functional.

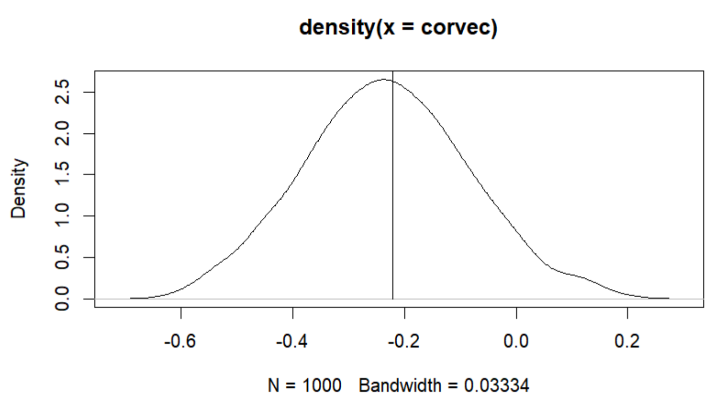
Freeze:



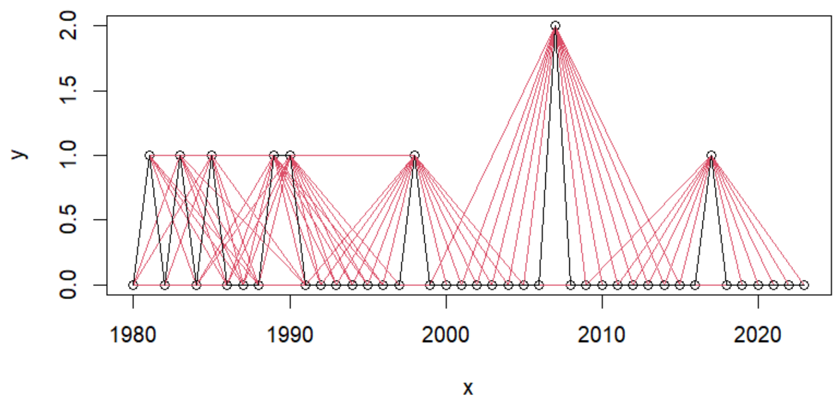
y
x
0.0
0.5
1.0
1.5
2.0
1980
1990
2000
2010
2020

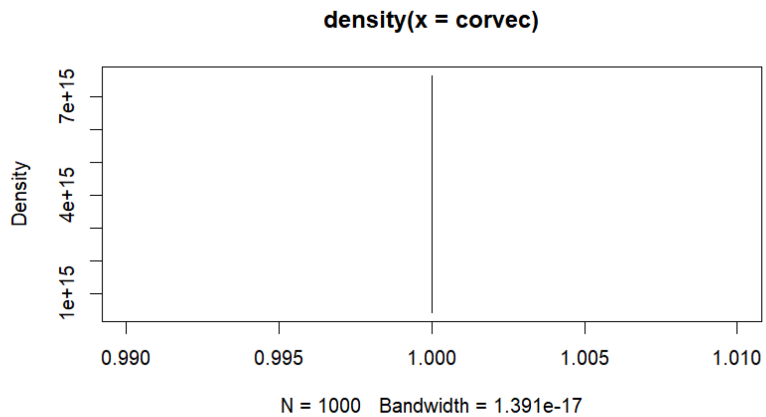


y
x
0.0
0.5
1.0
1.5
2.0
1980
1990
2000
2010
2020



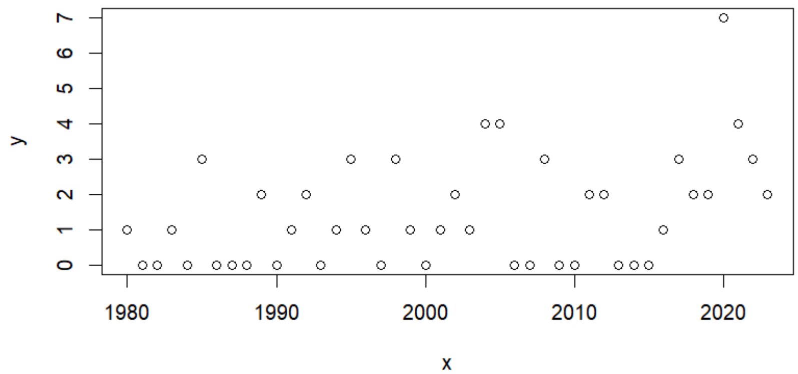
density(x = corvec)
Density
0.0
0.5
1.0
1.5
2.0
2.5
-0.6
-0.4
-0.2
0.0
0.2
N = 1000 Bandwidth = 0.03334

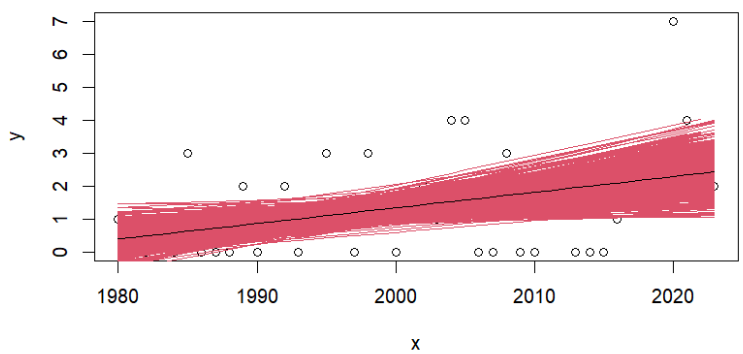


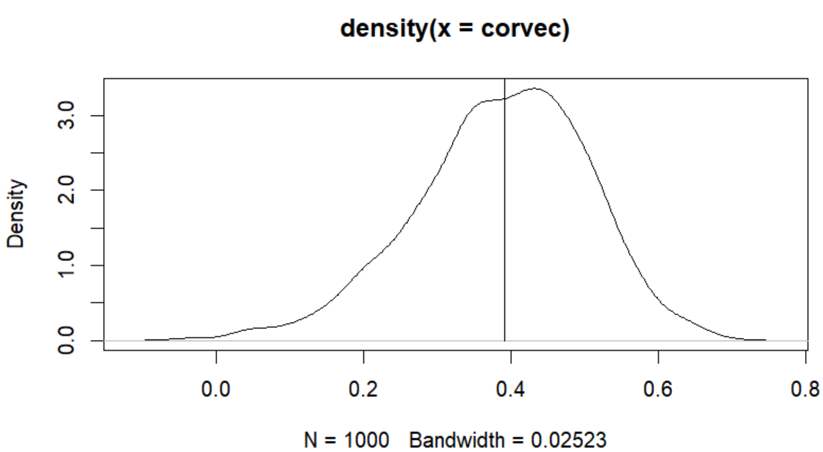


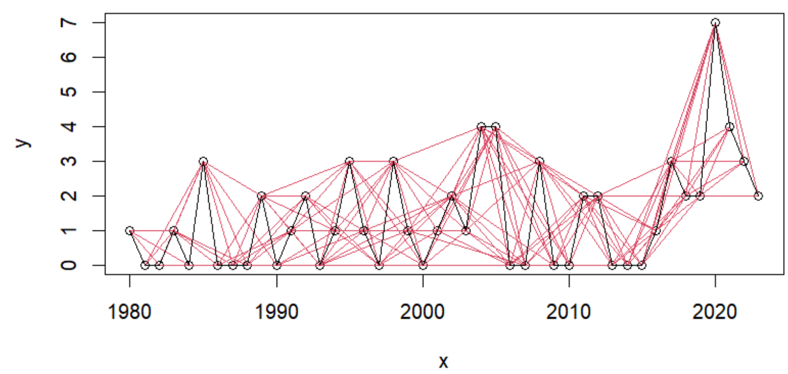
These plots show that the number of freezes in a year seems to have very little to no relation to time passing, and perhaps the frequency of freezes appears to go down slightly. But even that is a very slight trend, and any linear relationship between these two variables appears non-functional, as is evident by the correlation peak being around -0.2 and the lack of apparent patterns in the line chart with bootstrapped replicates.

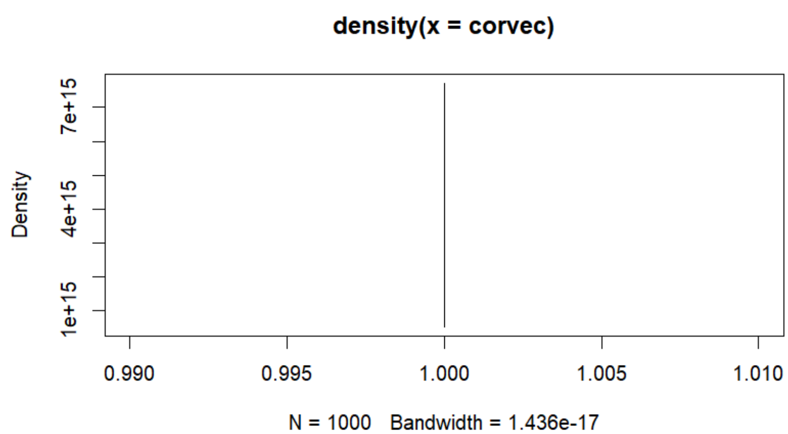
Tropical Cyclones:





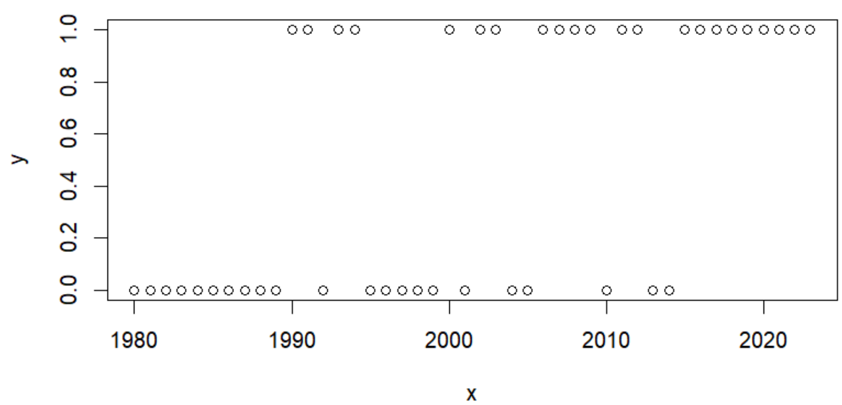




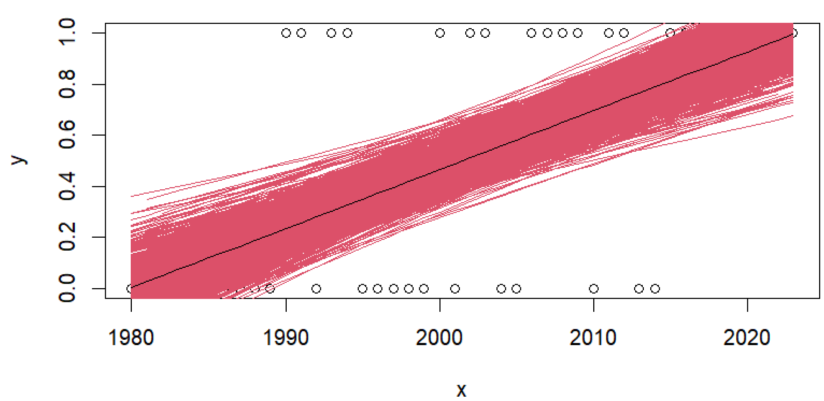


The scatter plots and the connected line graph with bootstrap replicates show that there is some relationship between the number of severe storms and time passing. However, the smoothed density plot for the correlation indicates that the correlation is centered around 0.4, which is very poor. Moreover, the linear regression plot with the bootstrap replicates shows that the relationship, if significant enough, is definitely not linear. Not many conclusions about the data can be drawn if we assume a linear relationship.

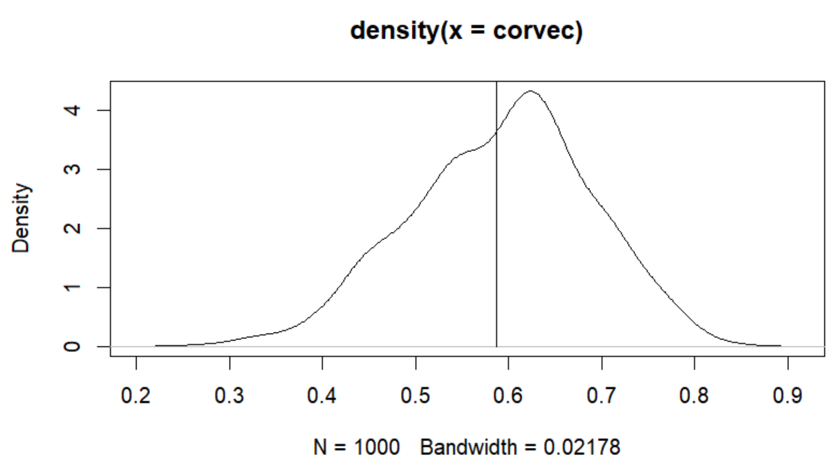
Wildfires:



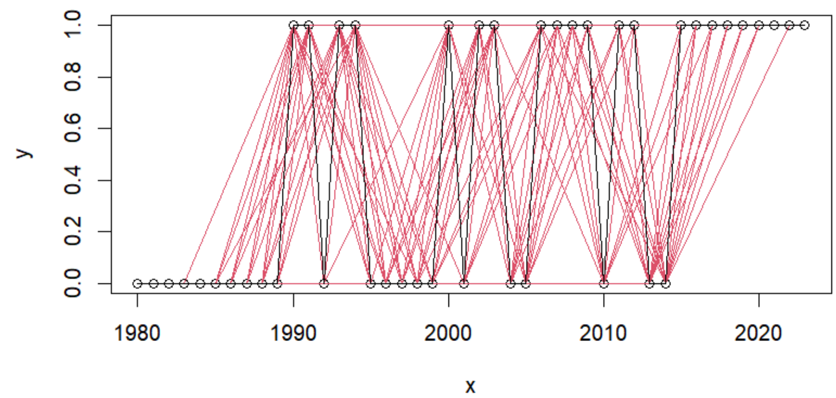
y
0.0
0.2
0.4
0.6
0.8
1.0
1980
1990
2000
2010
2020
x

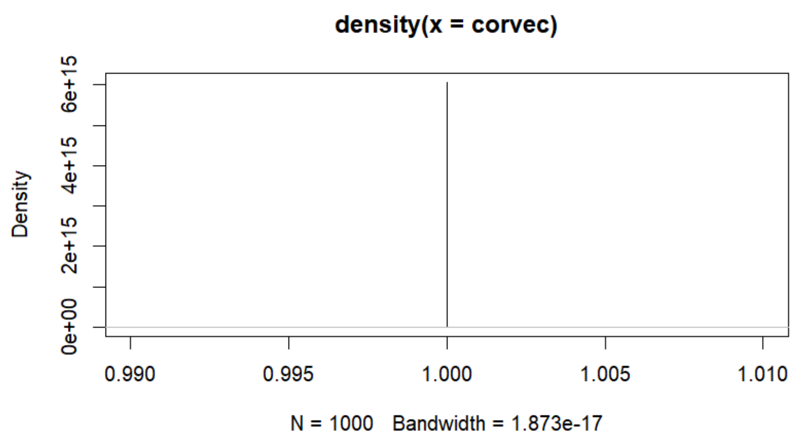


y
0.0
0.2
0.4
0.6
0.8
1.0
1980
1990
2000
2010
2020
x



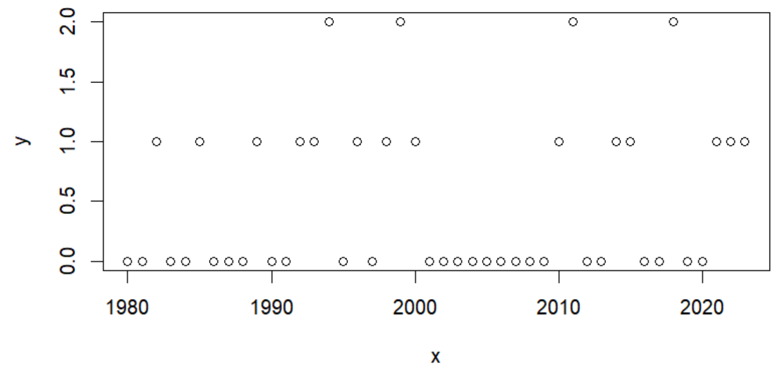
density(x = corvec)
Density
0
1
2
3
4
0.2
0.3
0.4
0.5
0.6
0.7
0.8
0.9
N = 1000 Bandwidth = 0.02178

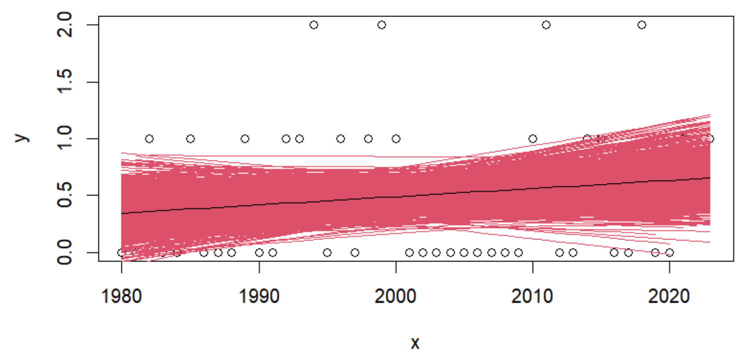


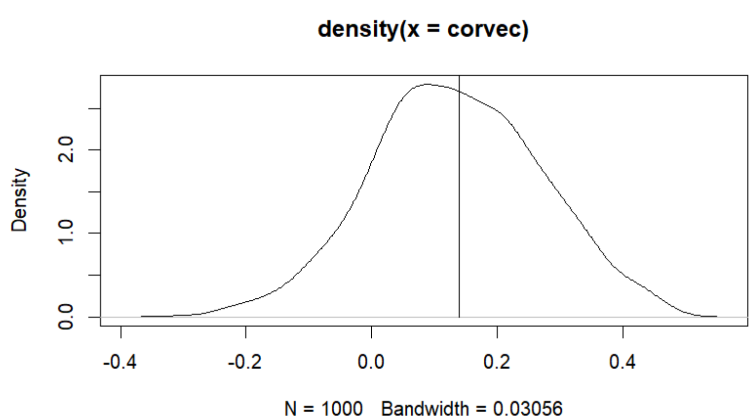


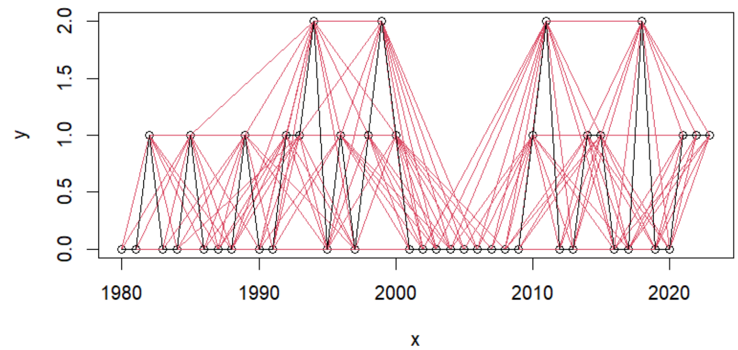
These plots reflect that the data seems to have two distinct values, 0 and 1, and it is challenging to establish a relationship between years passed and the number of Wildfires. When we look at the line chart with the bootstrapped replicates, there seems to be some relationship since wildfires seem to become more common as the years increase. However, it is clear from the plots that any linear relationship between these variables would be non-functional, especially considering the correlation peak seems to be around 0.6, which shows a somewhat weak linear relationship.

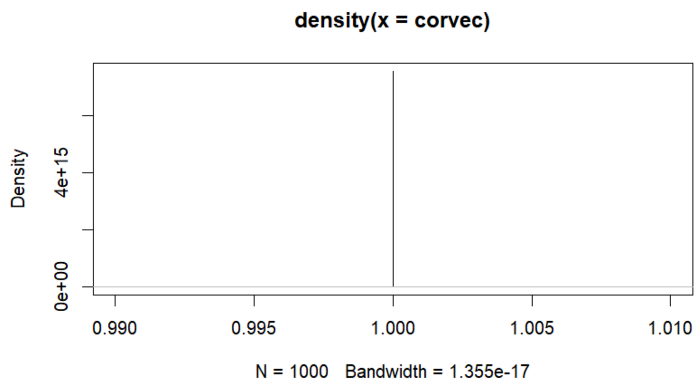
Winter Storms:





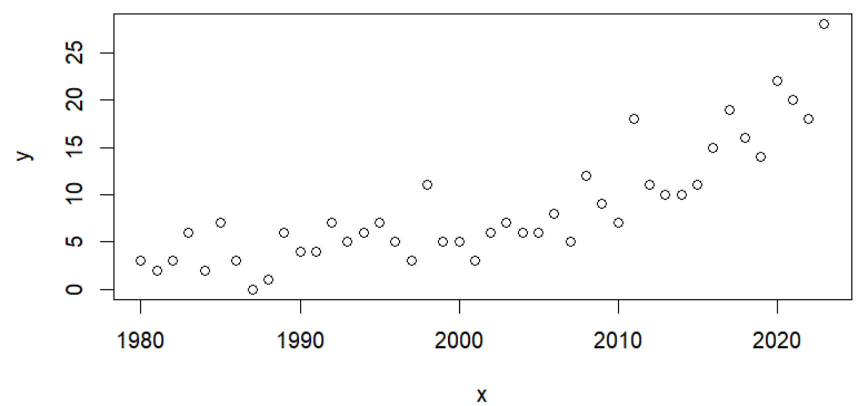


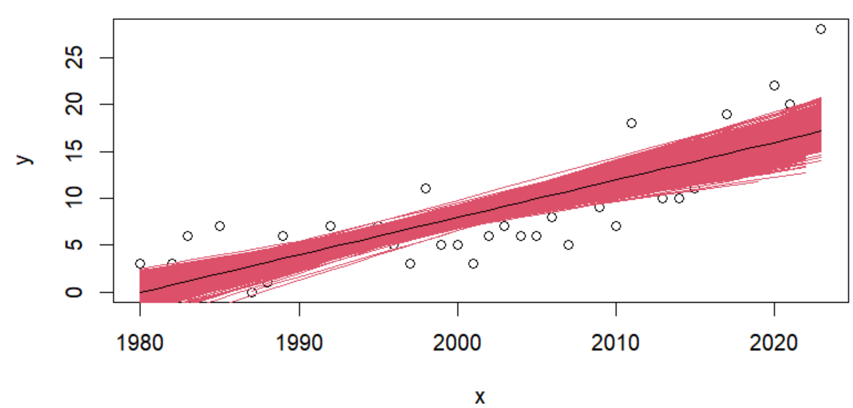


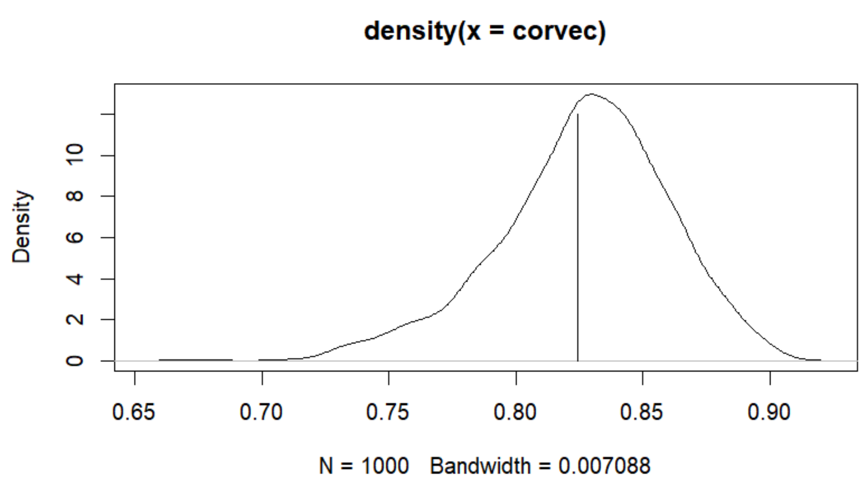


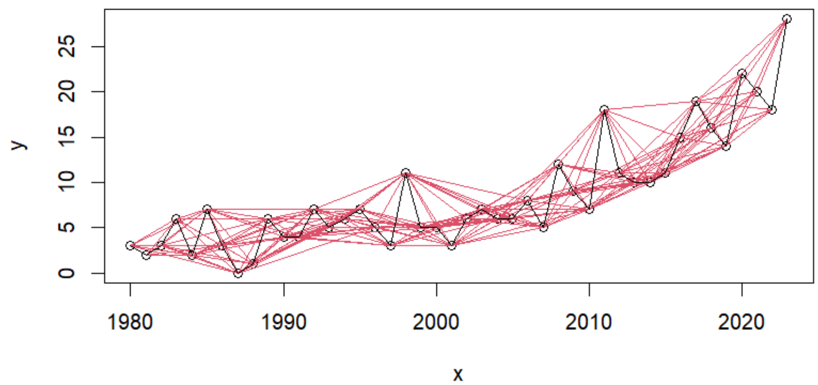
These plots indicate no relationship between years passing and the number of winter storms per year. While any linear relationship would certainly be non-functional, as evident from the lines of best-fit graph and the correlation peaking close to 0.1, any non-linear relationship also seems to be unlikely between these two variables. Suffice it to say that no clear relationship exists between these variables.

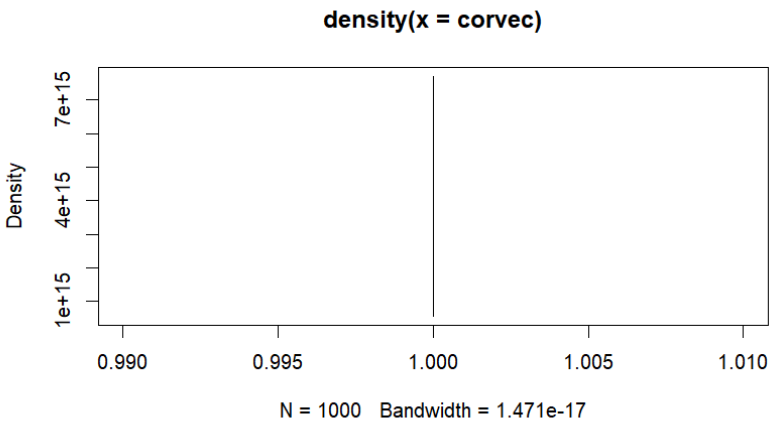
All disasters:







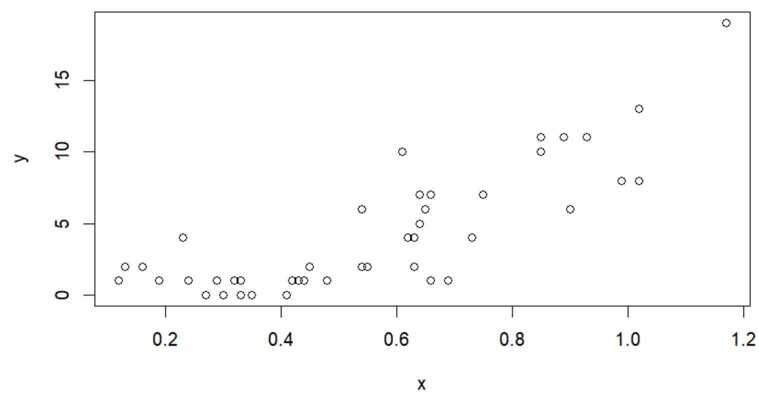


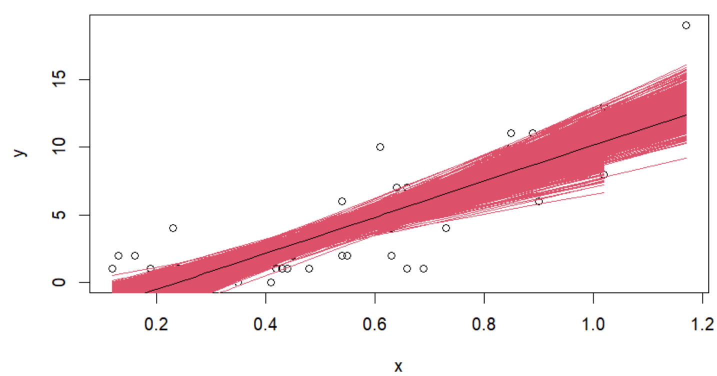


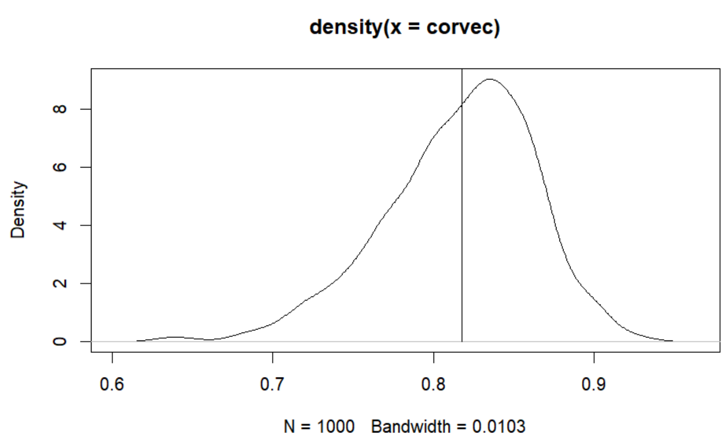
The scatter plot and the line chart with bootstrapped replicates show a very clear relationship between years and the number of total disasters in a year. There is a correlation here, but as is evident by the lines of best-fit plot and the correlation peak being somewhere between 0.8 and 0.85, the relationship is most likely non-linear. While establishing a linear relationship wouldn't produce completely inaccurate results, it would probably be somewhat non-functional, and a non-linear relationship might produce more accurate results.

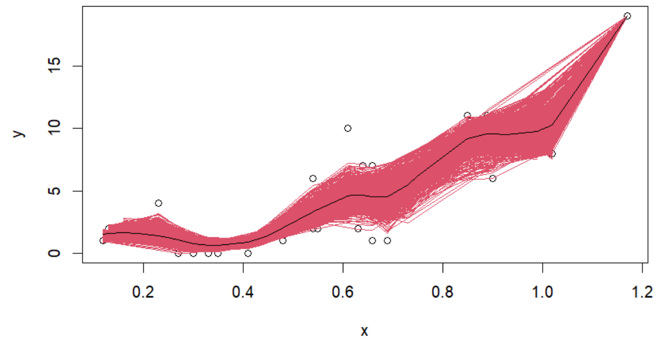
Comparing Disaster Statistics against Delta Time:

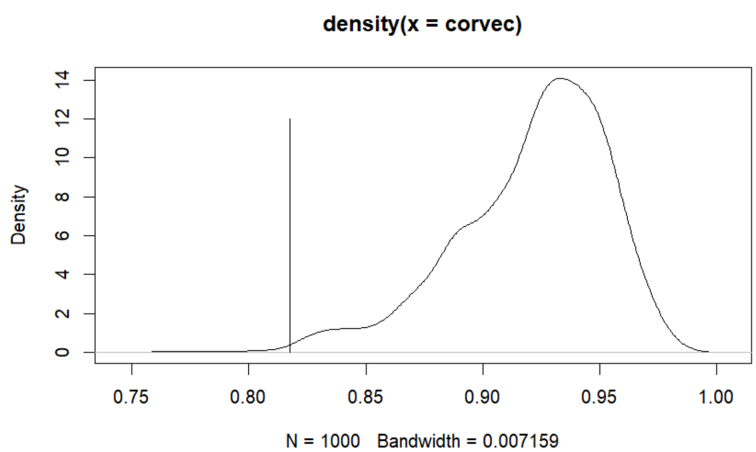
Severe Storms:





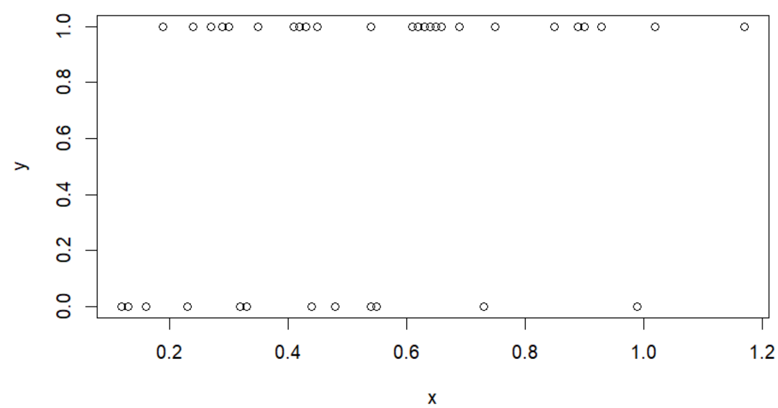




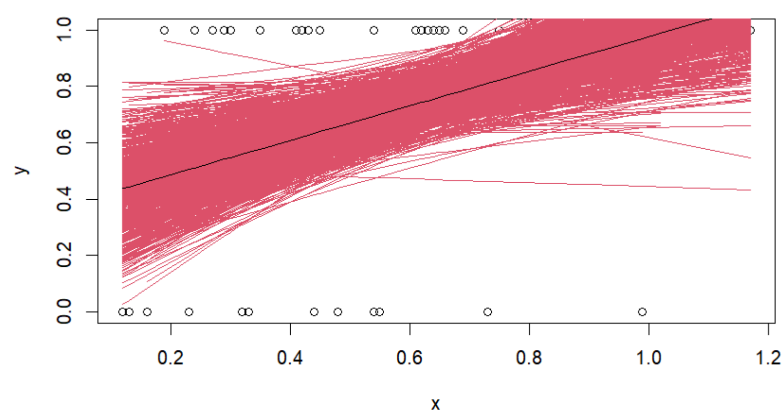


The results are very similar to when we compared with years passed. The scatter plots and the connected line graph with bootstrap replicates show that there is definitely a clear relationship between the number of severe storms and time passing. However, the smoothed density plot for the correlation shows that the correlation is centered around 0.82, which would capture much of the relationship but isn't always true. Moreover, the linear regression plot with the bootstrap replicates shows that while a linear relationship could be conceived and would be helpful, it won't always be accurate and that the relationship between these variables might be non-linear. Another piece of evidence for the relationship being very much present but non-linear is the correlation density peaks. The peak of the original data is around 0.82, but the peak of the bootstrapped samples is around 0.93-0.95. Both of these values are relatively high, meaning that a strong relationship between the variables definitely exists, but the discrepancy suggests that the relationship is probably non-linear.

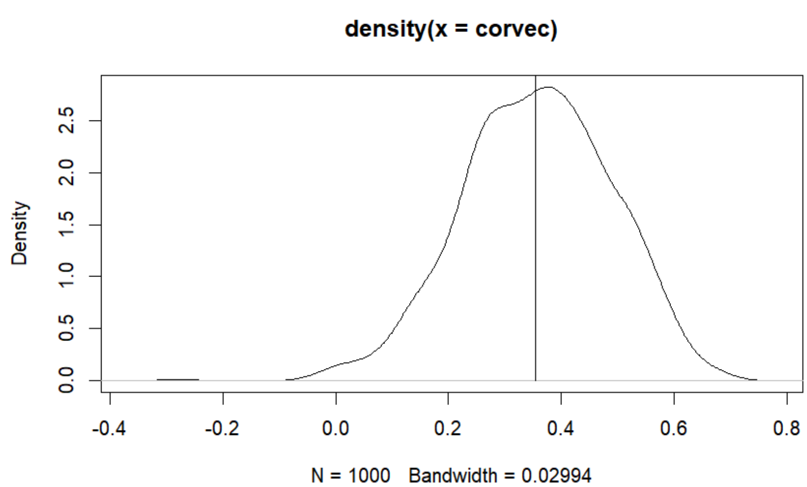
Droughts:



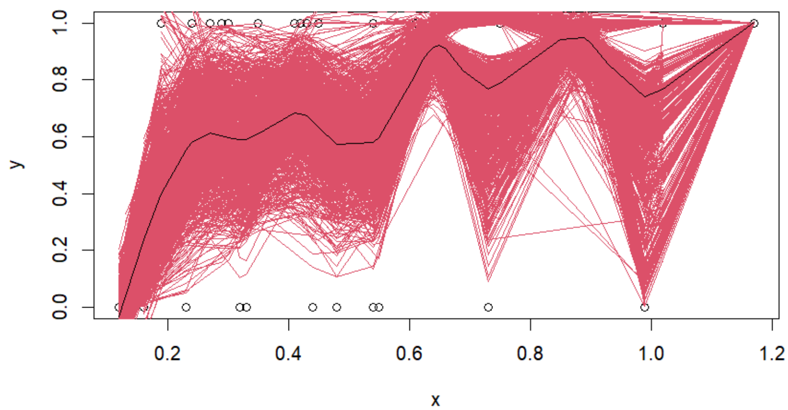
y
x
0.0
0.2
0.4
0.6
0.8
1.0
1.2

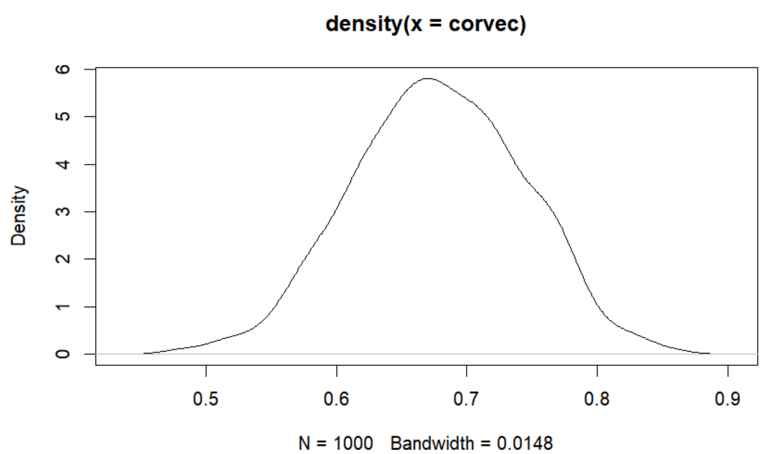


y
x
0.0
0.2
0.4
0.6
0.8
1.0
1.2



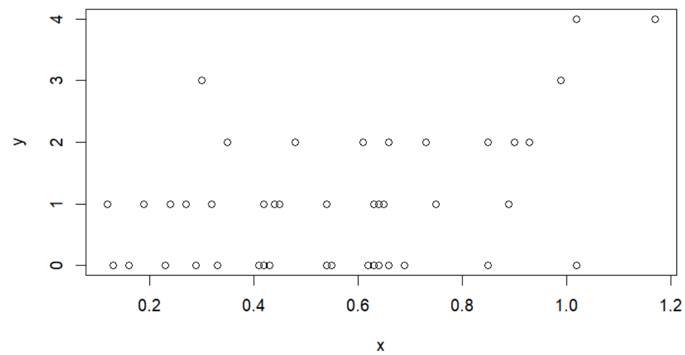
density(x = corvec)
Density
-0.4
-0.2
0.0
0.2
0.4
0.6
0.8
0.5
1.0
1.5
2.0
2.5
N = 1000 Bandwidth = 0.02994



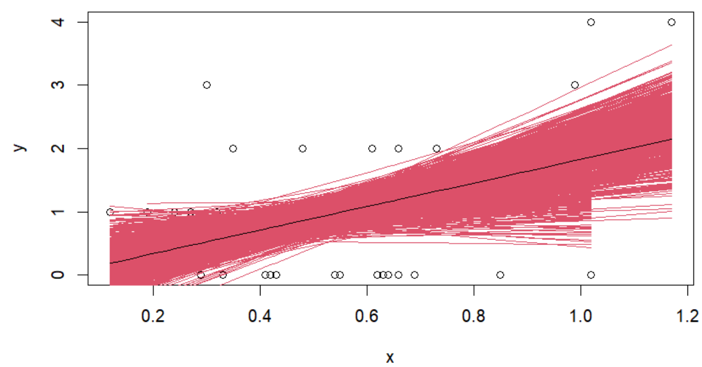


These results are almost identical to the ones we got when we compared with years passing. These plots reflect that the data seems to have two distinct values, 0 and 1, and that it is challenging to establish a relationship between years passed and the number of droughts. There does seem to be some relationship when we look at the line chart with the bootstrapped replicates since years without droughts seem to get more and more spaced out as the years increase. However, it is clear from the plots that any linear relationship between these variables would be non-functional, especially considering the correlation peak seems to be around 0.4, which shows a very weak linear relationship.

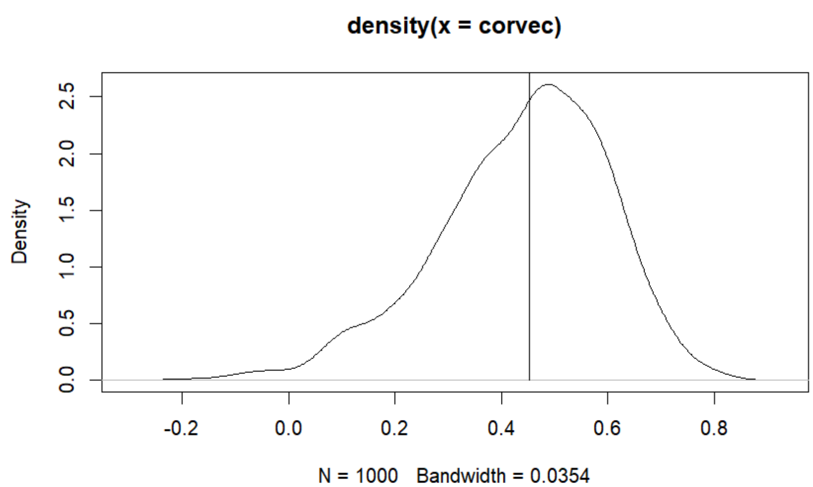
Floods:



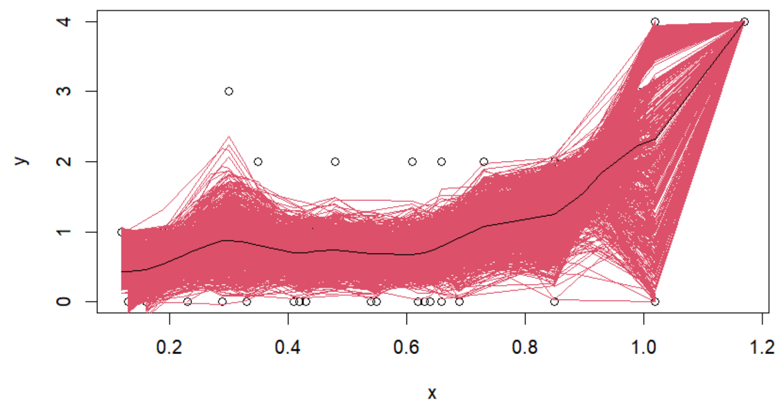
y
0
1
2
3
4
0.2
0.4
0.6
0.8
1.0
1.2
x

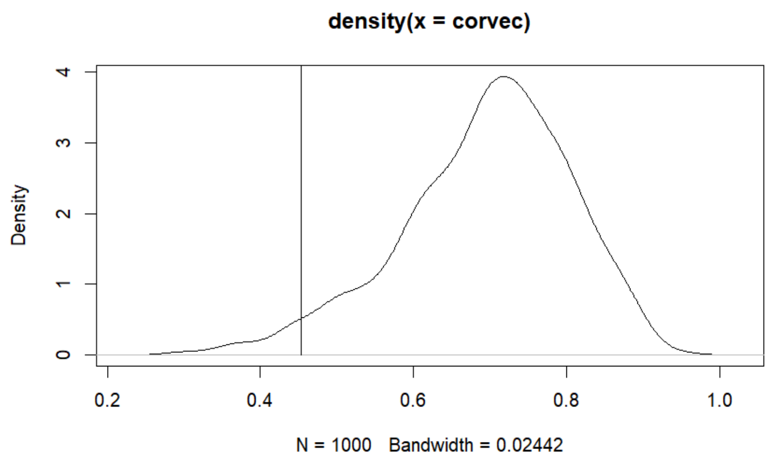


y
0
1
2
3
4
0.2
0.4
0.6
0.8
1.0
1.2
x



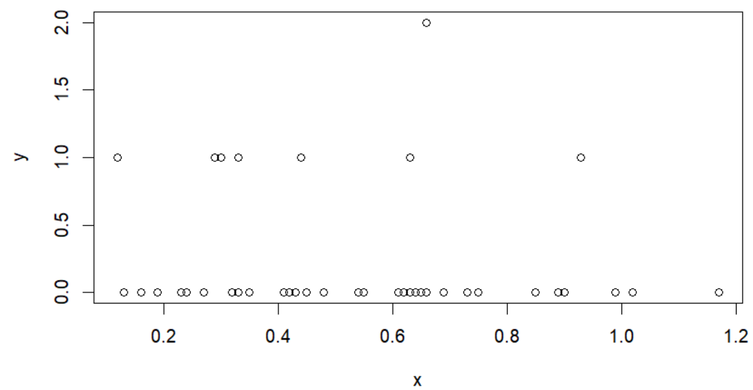
density(x = corvec)
Density
0.0
0.5
1.0
1.5
2.0
2.5
-0.2
0.0
0.2
0.4
0.6
0.8
N = 1000 Bandwidth = 0.0354



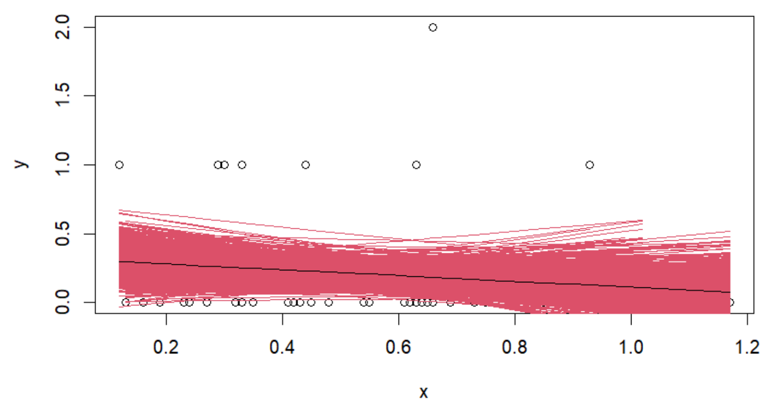


The scatter plot and the line chart show that the relationship between floods and time seems to exist but is not very strong. The relationship indicates that floods have generally increased, with one outlier between 1980 and 1990. However, the relationship very clearly does not seem linear, as is evident by the line of best-fit graph and the correlation plot, which appears to peak close to 0.4-0.5. Hence, a linear relationship would be non-functional. Another piece of evidence for the relationship being strictly non-linear(if it exists) would be that the correlation peak for the original dataset is somewhere around 0.4-0.5, but the peak for the bootstrapped samples is around 0.7, which is a huge gap, and almost certainly rules out the possibility of a linear relationship.

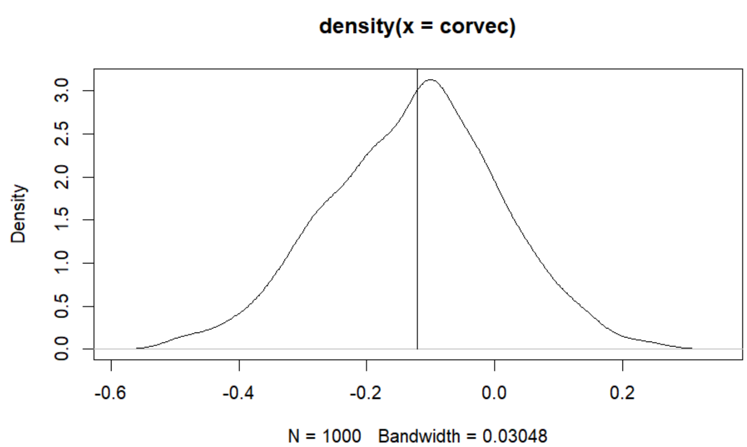
Freeze:



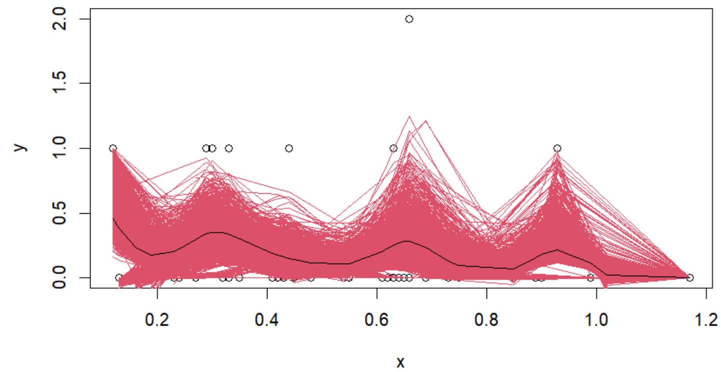
y
0.0
0.5
1.0
1.5
2.0
0.2
0.4
0.6
0.8
1.0
1.2
x

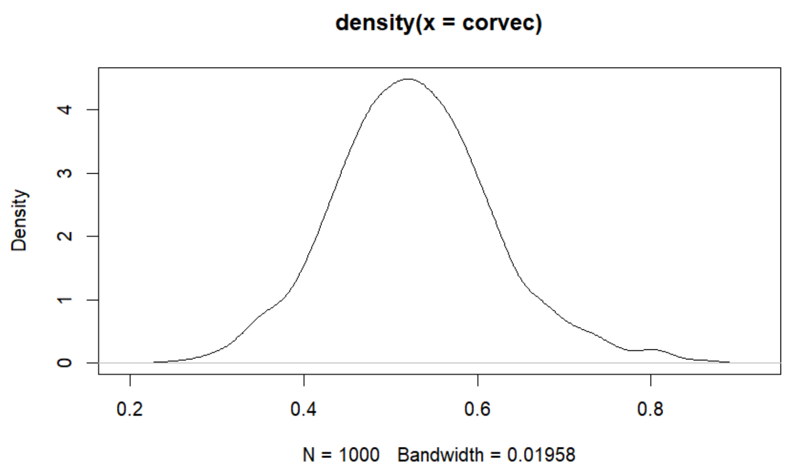


y
0.0
0.5
1.0
1.5
2.0
0.2
0.4
0.6
0.8
1.0
1.2
x



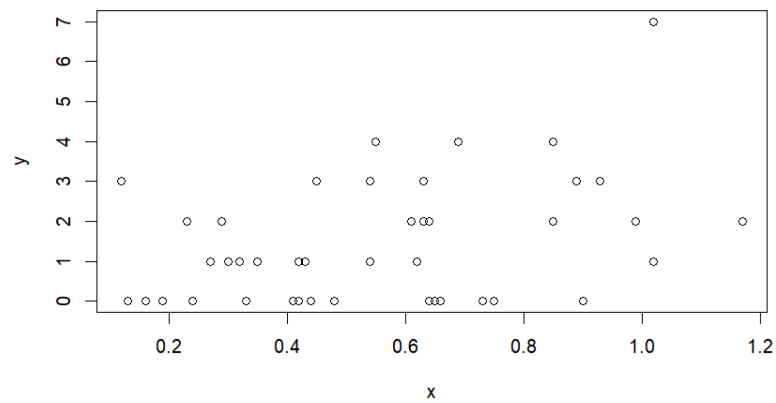
density(x = corvec)
Density
0.0
0.5
1.0
1.5
2.0
2.5
3.0
-0.6
-0.4
-0.2
0.0
0.2
N = 1000 Bandwidth = 0.03048

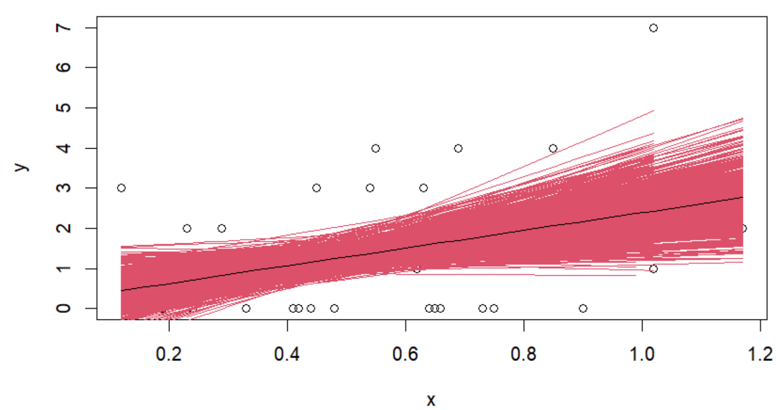


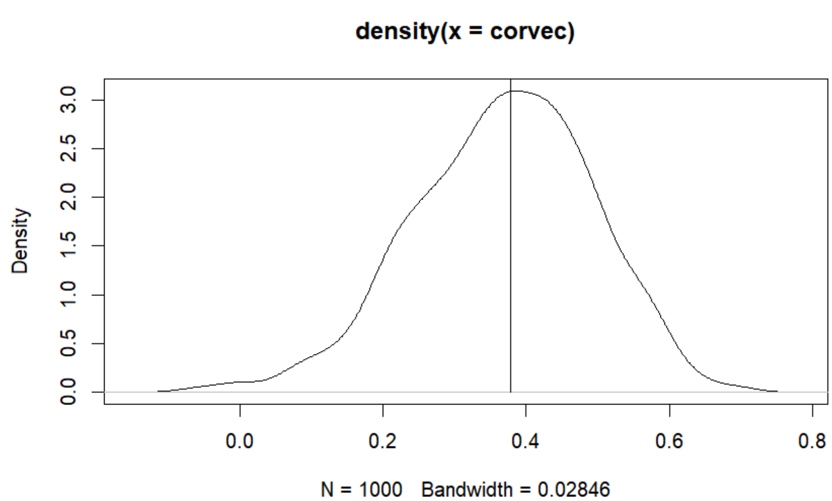


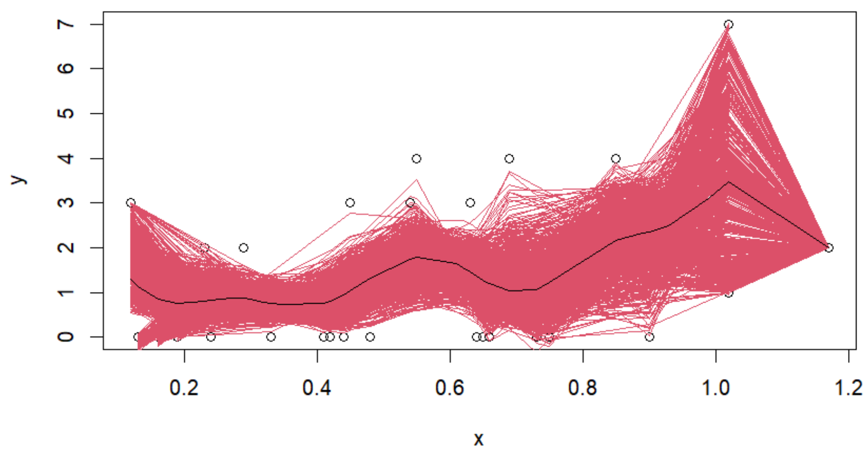
These plots show that the number of freezes in a year seems to have little to no relation to time passing, and perhaps the frequency of freezes appears to decrease slightly. But even that is a very slight trend, and any linear relationship between these two variables appears non-functional, as is evident by the correlation peak being around -0.1 and the lack of apparent patterns in the line chart with bootstrapped replicates. Even if a slight relationship were to exist, it most certainly isn’t linear since the correlation peak of the bootstrapped replicates is far off(around 0.5) from the peak of the original peak.

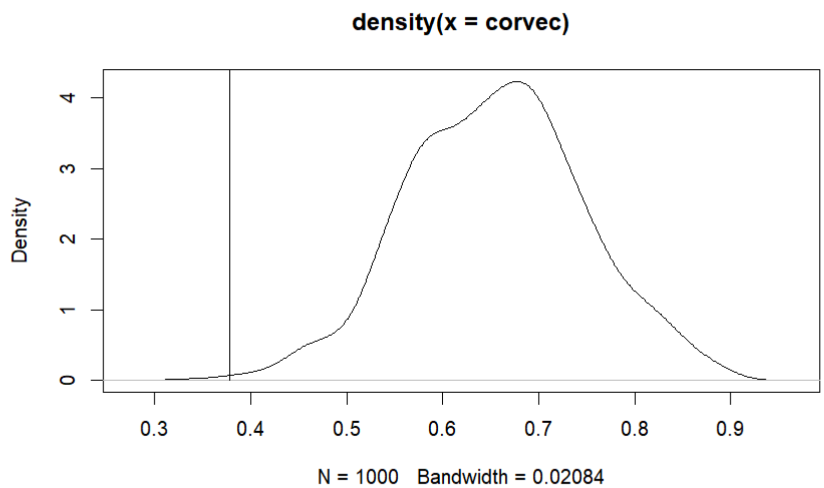
Tropical Cyclones:





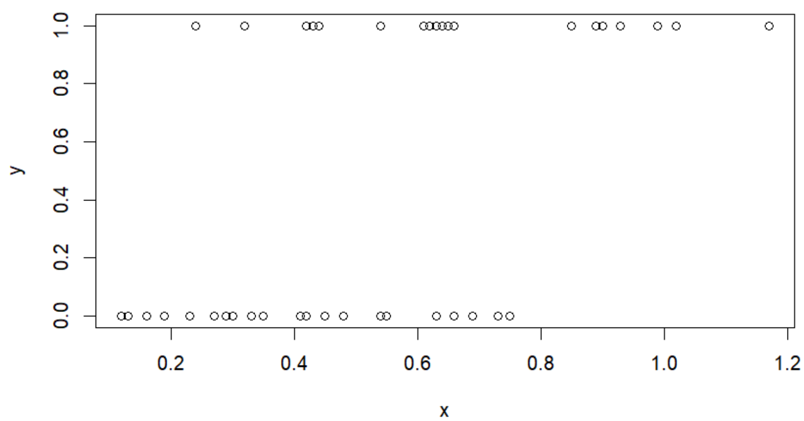




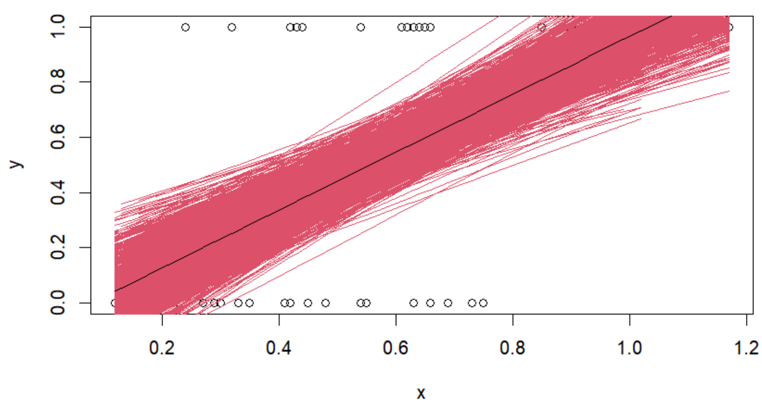


Very similar to the results we got when we compared with years passed, the scatter plots and the connected line graph with bootstrap replicates show that there is some relationship between the number of severe storms and time passing. However, the smoothed density plot for the correlation indicates that the correlation is centered around 0.4, which is very poor. Moreover, the linear regression plot with the bootstrap replicates shows that the relationship, if significant enough, is definitely not linear. The fact that the correlation peak of the bootstrapped samples is very far from the original data's peak(around 0.7) further suggests a non-linear relationship. Not many conclusions about the data can be drawn if we assume a linear relationship.

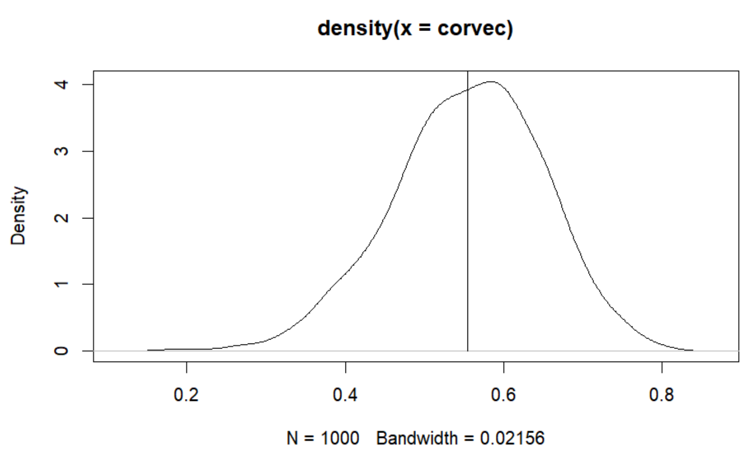
Wildfires:



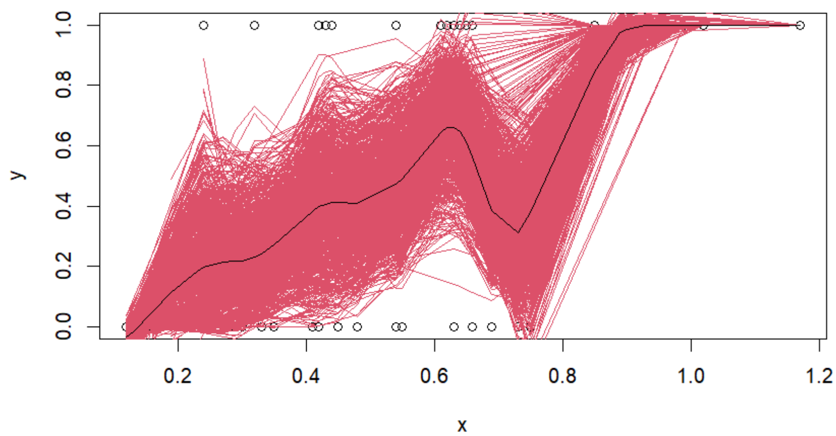
y
0.0
0.2
0.4
0.6
0.8
1.0
0.2
0.4
0.6
0.8
1.0
1.2
x

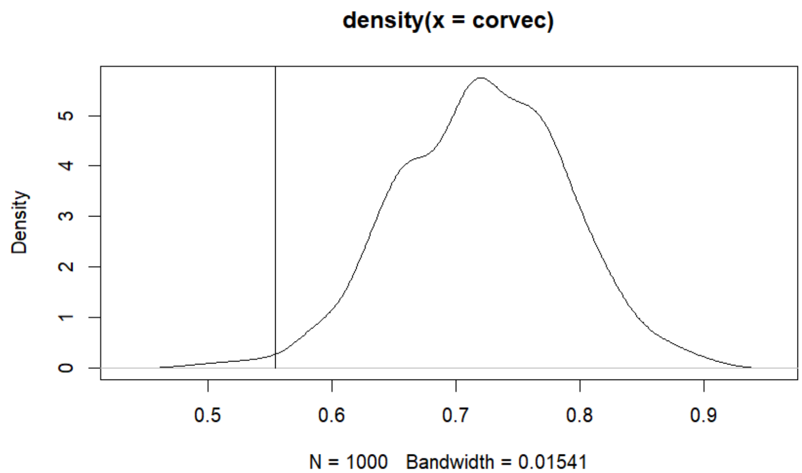


y
0.0
0.2
0.4
0.6
0.8
1.0
0.2
0.4
0.6
0.8
1.0
1.2
x



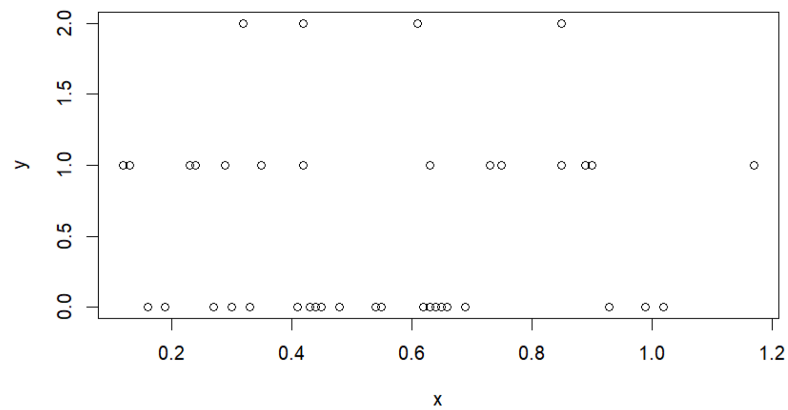
density(x = corvec)
Density
0
1
2
3
4
0.2
0.4
0.6
0.8
N = 1000 Bandwidth = 0.02156

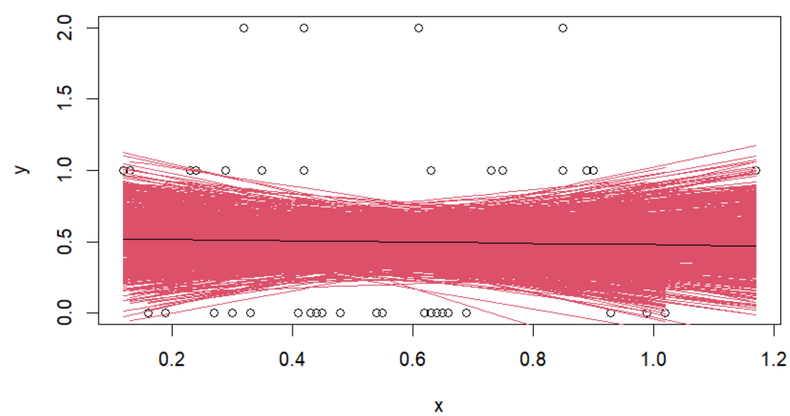


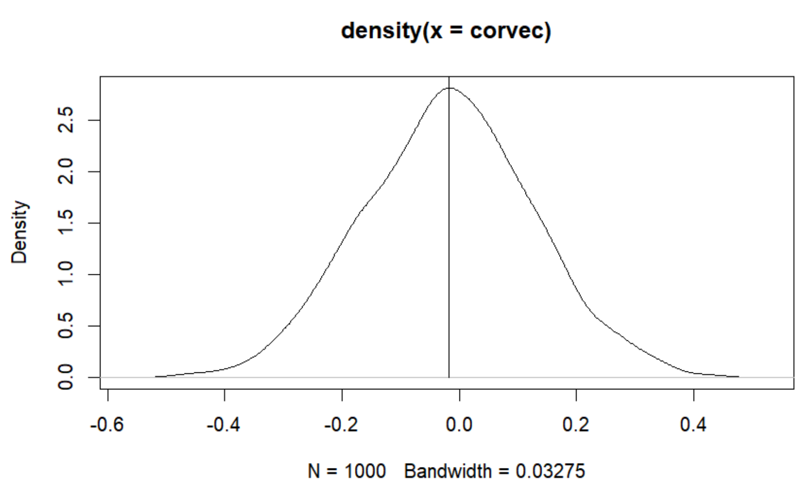


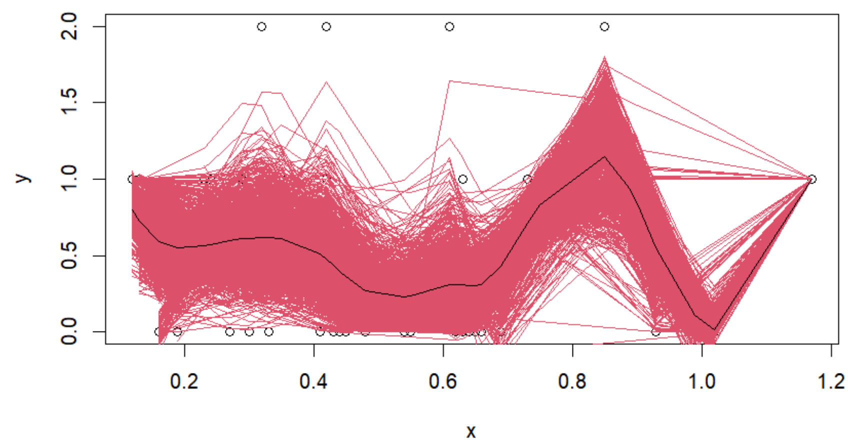
Again, very similar to comparisons with the years passing, these plots reflect that the data seems to have two distinct values, 0 and 1, and it is challenging to establish a relationship between years passed and the number of Wildfires. When we look at the line chart with the bootstrapped replicates, there seems to be some relationship since wildfires seem to become more common as the years increase. However, it is clear from the plots that any linear relationship between these variables would be non-functional, especially considering the correlation peak for the original data seems to be around 0.55, and the peak for the bootstrapped samples seems to be around 0.7, which shows that the relationship(if significant enough) is most likely non-linear.

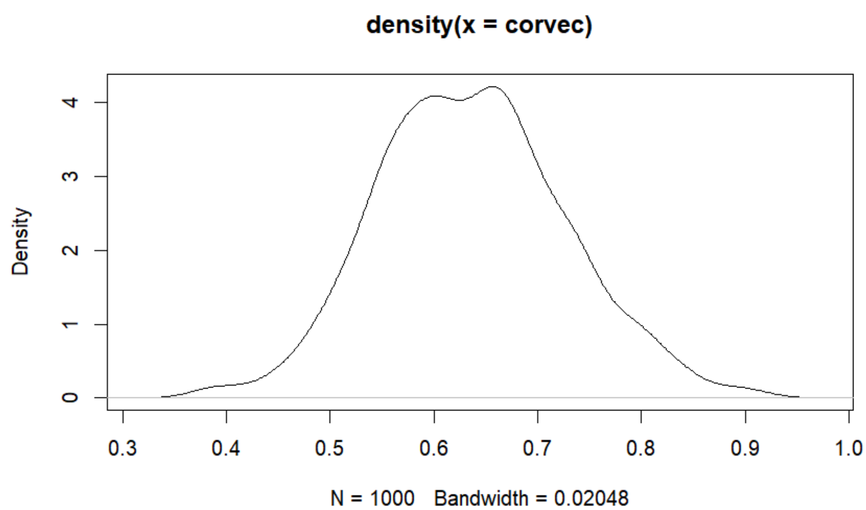
Winter Storms:





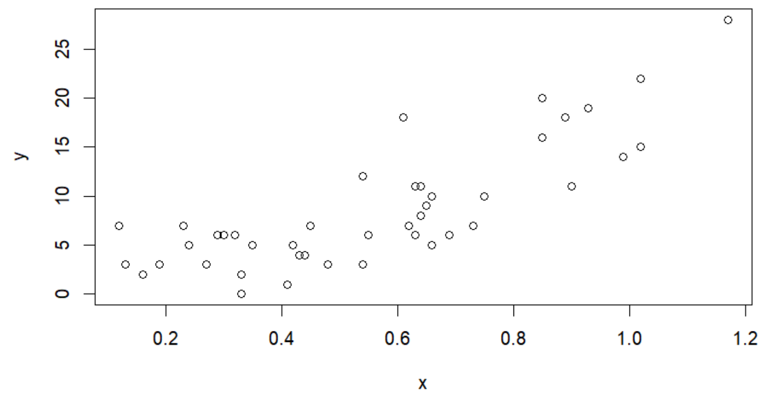


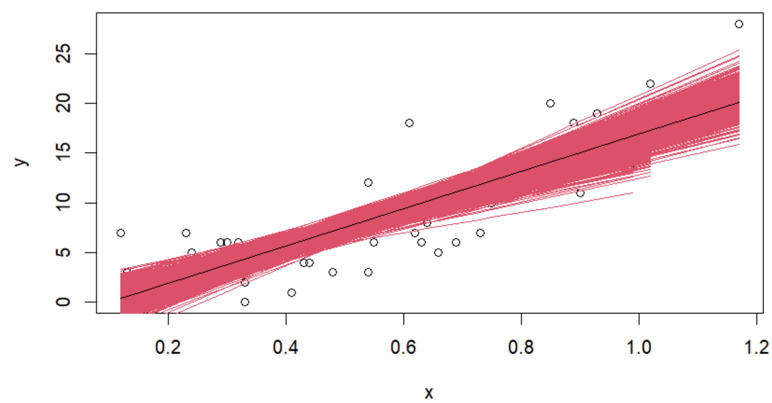


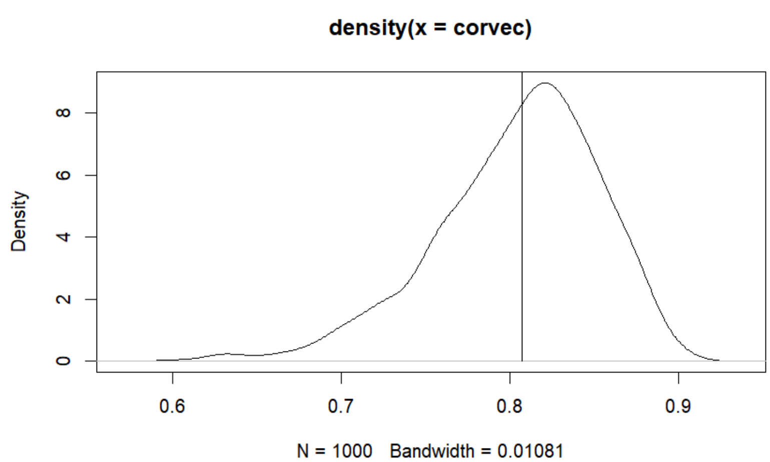


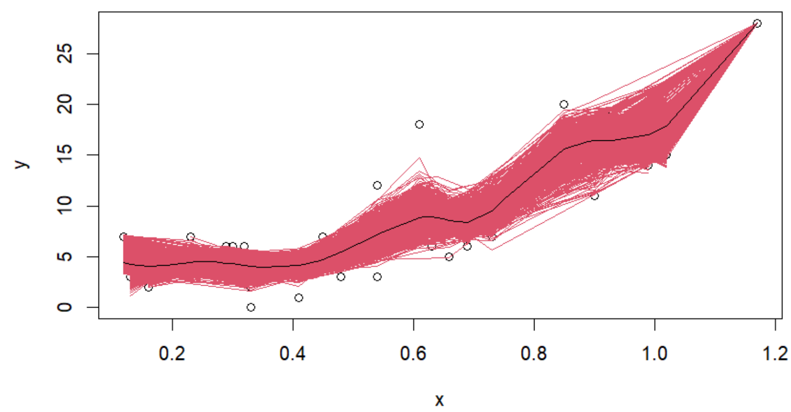
These plots indicate no relationship between years passing and the number of winter storms per year. While any linear relationship would certainly be non-functional, as evident from the lines of best-fit graph and the correlation peaking close to 0, any non-linear relationship also seems to be absent between these two variables, and the peak for the bootstrapped samples is randomly around 0.6. Suffice it to say that no clear relationship exists between these variables, and if a slight relationship exists, then it is definitely non-linear.

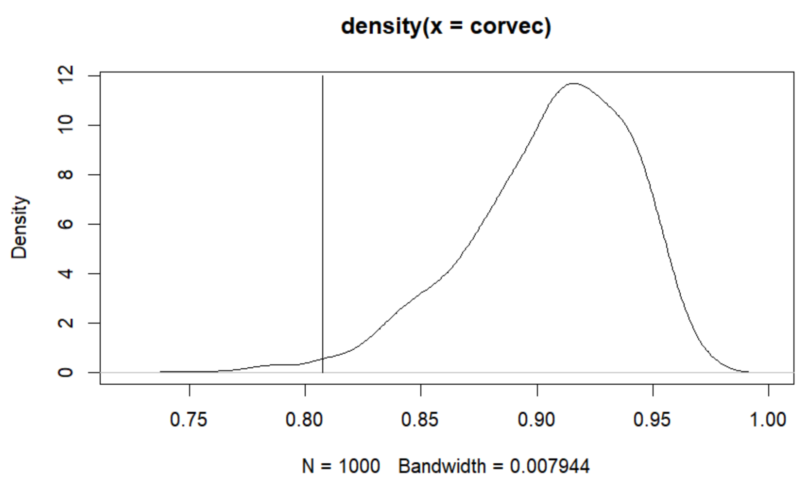
All disasters:







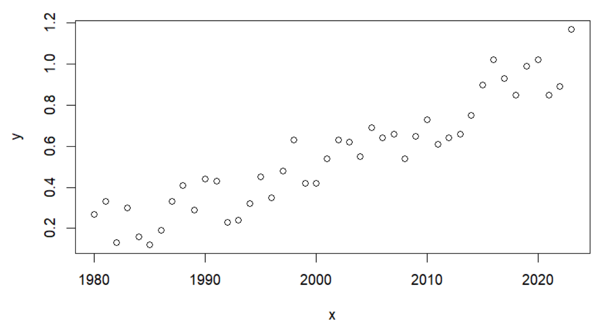


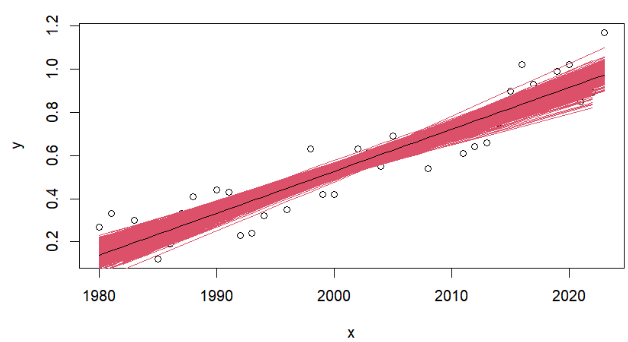


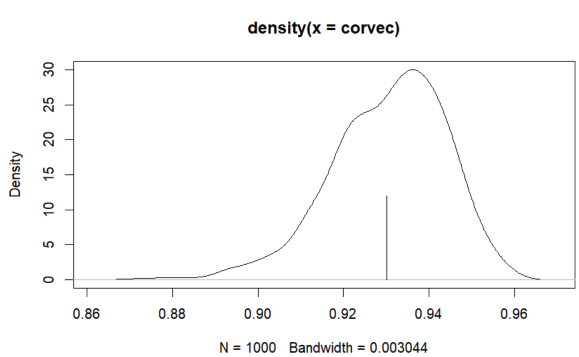
Again, very similar to the comparison with years passing, the scatter plot and the line chart with bootstrapped replicates show a very clear relationship between years and the number of total disasters in a year. There is a correlation here, but as is evident by the lines of best-fit plot and the correlation peak being somewhere close to 0.8, the relationship is somewhat accurate when we assume that it is linear. While establishing a linear relationship wouldn't produce completely inaccurate results, it would probably be somewhat non-functional, and a non-linear relationship might yield more accurate results as is indicated by the fact that the correlation peak of the bootstrapped replicates is close to 0.92 which is significantly higher than the peak of the original data, making it likely that the true nature of the relationship between these variables seems to be non-linear in nature.

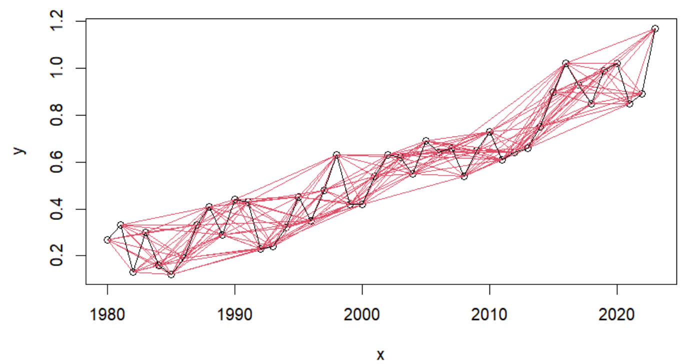
Conclusion:

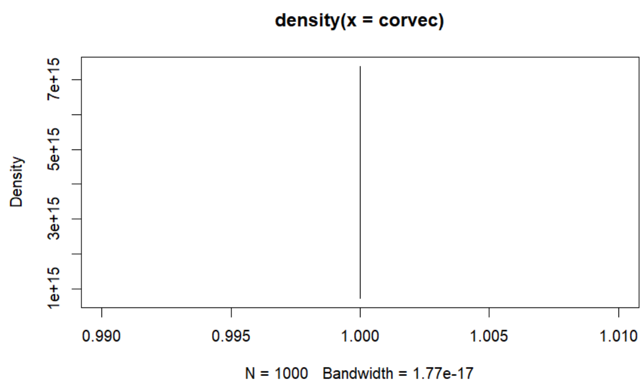
It seems that a lot of disasters do have significant trends when compared with time and change in temperature. Some disasters have slight trends, while others seem to bear no connection to these variables. One apparent conclusion that can be drawn is that the trends seem to be very similar when we compare the frequency of a disaster with time or with change in temperature. This can be explained when we look at the fact the time and change in temperature do have a clear linear relationship with each other.











The scatter plot and the plot of the lines of best-fit show that there is a clear linear relationship between time and change in temperature. Furthermore, the line chart with the bootstrapped replicates shows little variation between the bootstrapped replicates and the original data. The correlation peaks of the original data and the bootstrapped replicates are also both relatively high(both somewhere between 0.93 and 0.94) which means that the relationship is very strong and that it is most likely linear. This would explain why our results seem to be very similar for disasters when we compare them to time or with change in temperature per year.

Code in R:

```
# Load necessary libraries
library(tidyverse)
library(moderndive)
library(skimr)
library(corrplot)
```

```
# Load the provided data/functions
source("D:/Uni Resources/Stats/assignment1/fixregday1.pck")

# Load the dataset
data <- read.csv("D:/Uni Resources/Stats/assignment1/d.csv")


head(data)
str(data)
summary(data)
skim(data)

#storm
bootmatsmooth(NOAAGISSWD$Year, NOAAGISSWD$Severe.Storm.Count)
bootmatlin(NOAAGISSWD$Year, NOAAGISSWD$Severe.Storm.Count)

#drought
bootmatsmooth(NOAAGISSWD$Year, NOAAGISSWD$Drought.Count)
bootmatlin(NOAAGISSWD$Year, NOAAGISSWD$Drought.Count)

#floods
bootmatsmooth(NOAAGISSWD$Year, NOAAGISSWD$Flooding.Count)
bootmatlin(NOAAGISSWD$Year, NOAAGISSWD$Flooding.Count)

#freeze
bootmatsmooth(NOAAGISSWD$Year, NOAAGISSWD$Freeze.Count)
bootmatlin(NOAAGISSWD$Year, NOAAGISSWD$Freeze.Count)

#Tropical Cylones
bootmatsmooth(NOAAGISSWD$Year, NOAAGISSWD$Tropical.Cyclone.Count)
bootmatlin(NOAAGISSWD$Year, NOAAGISSWD$Tropical.Cyclone.Count)

#Wildfires
bootmatsmooth(NOAAGISSWD$Year, NOAAGISSWD$Wildfire.Count)
bootmatlin(NOAAGISSWD$Year, NOAAGISSWD$Wildfire.Count)

#Winter Storms
bootmatsmooth(NOAAGISSWD$Year, NOAAGISSWD$Winter.Storm.Count)
bootmatlin(NOAAGISSWD$Year, NOAAGISSWD$Winter.Storm.Count)

#All disasters
bootmatsmooth(NOAAGISSWD$Year, NOAAGISSWD$All.Disasters.Count)
bootmatlin(NOAAGISSWD$Year, NOAAGISSWD$All.Disasters.Count)
```

```
#-----------------------------------------------------------------------------------------
#Comparing with delta temperature

#storm
bootmatsmooth(NOAAGISSWD$delta.temp, NOAAGISSWD$Severe.Storm.Count)
bootmatlin(NOAAGISSWD$delta.temp, NOAAGISSWD$Severe.Storm.Count)

#drought
bootmatsmooth(NOAAGISSWD$delta.temp, NOAAGISSWD$Drought.Count)
bootmatlin(NOAAGISSWD$delta.temp, NOAAGISSWD$Drought.Count)

#floods
bootmatsmooth(NOAAGISSWD$delta.temp, NOAAGISSWD$Flooding.Count)
bootmatlin(NOAAGISSWD$delta.temp, NOAAGISSWD$Flooding.Count)

#freeze
bootmatsmooth(NOAAGISSWD$delta.temp, NOAAGISSWD$Freeze.Count)
bootmatlin(NOAAGISSWD$delta.temp, NOAAGISSWD$Freeze.Count)

#Tropical Cylones
bootmatsmooth(NOAAGISSWD$delta.temp, NOAAGISSWD$Tropical.Cyclone.Count)
bootmatlin(NOAAGISSWD$delta.temp, NOAAGISSWD$Tropical.Cyclone.Count)

#Wildfires
bootmatsmooth(NOAAGISSWD$delta.temp, NOAAGISSWD$Wildfire.Count)
bootmatlin(NOAAGISSWD$delta.temp, NOAAGISSWD$Wildfire.Count)

#Winter Storms
bootmatsmooth(NOAAGISSWD$delta.temp, NOAAGISSWD$Winter.Storm.Count)
bootmatlin(NOAAGISSWD$delta.temp, NOAAGISSWD$Winter.Storm.Count)

#All disasters
bootmatsmooth(NOAAGISSWD$delta.temp, NOAAGISSWD$All.Disasters.Count)
bootmatlin(NOAAGISSWD$delta.temp, NOAAGISSWD$All.Disasters.Count)


#years and delta temp

bootmatsmooth(NOAAGISSWD$Year, NOAAGISSWD$delta.temp)
bootmatlin(NOAAGISSWD$Year, NOAAGISSWD$delta.temp)
```