Unemployment Report

Jeff Krueger, Corey Lund, Fred Mwangi, and Nemanja Orescanin

11/29/2020

library(devtools)

## Loading required package: usethis

library(blsAPI)  
library(rjson)  
library(curl)  
library(RCurl)  
library(knitr)  
library(readr)

##   
## Attaching package: 'readr'

## The following object is masked from 'package:curl':  
##   
## parse\_date

library(dplyr)

##   
## Attaching package: 'dplyr'

## The following objects are masked from 'package:stats':  
##   
## filter, lag

## The following objects are masked from 'package:base':  
##   
## intersect, setdiff, setequal, union

library(tidyverse)

## -- Attaching packages --------------------------------------- tidyverse 1.3.0 --

## v ggplot2 3.3.2 v purrr 0.3.4  
## v tibble 3.0.4 v stringr 1.4.0  
## v tidyr 1.1.2 v forcats 0.5.0

## -- Conflicts ------------------------------------------ tidyverse\_conflicts() --  
## x tidyr::complete() masks RCurl::complete()  
## x dplyr::filter() masks stats::filter()  
## x dplyr::lag() masks stats::lag()  
## x readr::parse\_date() masks curl::parse\_date()

library(rmarkdown)  
library(pastecs)

##   
## Attaching package: 'pastecs'

## The following object is masked from 'package:tidyr':  
##   
## extract

## The following objects are masked from 'package:dplyr':  
##   
## first, last

library(ggplot2)  
library(corrplot)

## corrplot 0.84 loaded

library(lmtest)

## Loading required package: zoo

##   
## Attaching package: 'zoo'

## The following objects are masked from 'package:base':  
##   
## as.Date, as.Date.numeric

##   
## Attaching package: 'lmtest'

## The following object is masked from 'package:RCurl':  
##   
## reset

library(forecast)

## Registered S3 method overwritten by 'quantmod':  
## method from  
## as.zoo.data.frame zoo

#loads the necessary packages

First we need to load the libraries that we will be using into our report. The above code should load the proper libraries. R Studio will prompt you to install the packages if necessary.

merged\_final <- read\_csv("csv files/mergedFinal.csv")

##   
## -- Column specification --------------------------------------------------------  
## cols(  
## year = col\_double(),  
## period = col\_character(),  
## periodName = col\_character(),  
## value = col\_double(),  
## state = col\_character(),  
## quarter = col\_double(),  
## annual = col\_double(),  
## HPI = col\_double(),  
## MHI = col\_double(),  
## RMHI = col\_double(),  
## poverty = col\_double(),  
## population = col\_double(),  
## sp500 = col\_double(),  
## log\_pop = col\_double(),  
## log\_RMHI = col\_double()  
## )

#Reads final file into merged dataframe  
  
merged\_final$log\_pop = as.numeric(log(merged\_final$population))  
  
merged\_final$log\_RMHI = as.numeric(log(merged\_final$RMHI))

As a final step in scrubbing our data, we created new variables to log the population and median household income variables and store it in our dataframe. By taking the logarithmic values, these variables become more comparable to the existing variables we have within our dataset.

Below you can see the table output once all of our data has been merged.

kable(merged\_final[1:6,], caption = "Table Including All Variables")

Table Including All Variables

|  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| year | period | periodName | value | state | quarter | annual | HPI | MHI | RMHI | poverty | population | sp500 | log\_pop | log\_RMHI |
| 1984 | M12 | December | 9.3 | AK | 4 | 1 | 153.82 | 32356 | 74689 | 9.6 | 513702 | 1.40 | 13.14940 | 11.22109 |
| 1985 | M12 | December | 10.1 | AK | 4 | 1 | 143.07 | 34782 | 77625 | 8.7 | 532495 | 26.33 | 13.18533 | 11.25964 |
| 1986 | M12 | December | 11.0 | AK | 4 | 1 | 138.24 | 31356 | 68775 | 11.4 | 544268 | 14.62 | 13.20720 | 11.13860 |
| 1987 | M12 | December | 9.5 | AK | 4 | 1 | 106.02 | 33233 | 70468 | 12.0 | 539309 | 2.03 | 13.19804 | 11.16291 |
| 1988 | M12 | December | 8.0 | AK | 4 | 1 | 133.16 | 33103 | 67745 | 11.0 | 541983 | 12.40 | 13.20299 | 11.12351 |
| 1989 | M12 | December | 7.1 | AK | 4 | 1 | 102.92 | 36006 | 70599 | 10.5 | 547159 | 27.25 | 13.21249 | 11.16477 |

#Descriptive Statistics  
stat.desc(merged\_final)

## year period periodName value state quarter annual  
## nbr.val 1.800000e+03 NA NA 1.800000e+03 NA 1800 1800  
## nbr.null 0.000000e+00 NA NA 0.000000e+00 NA 0 0  
## nbr.na 0.000000e+00 NA NA 0.000000e+00 NA 0 0  
## min 1.984000e+03 NA NA 1.800000e+00 NA 4 1  
## max 2.018000e+03 NA NA 1.440000e+01 NA 4 1  
## range 3.400000e+01 NA NA 1.260000e+01 NA 0 0  
## sum 3.602250e+06 NA NA 9.793800e+03 NA 7200 1800  
## median 2.001500e+03 NA NA 5.100000e+00 NA 4 1  
## mean 2.001250e+03 NA NA 5.441000e+00 NA 4 1  
## SE.mean 2.373587e-01 NA NA 4.648803e-02 NA 0 0  
## CI.mean.0.95 4.655278e-01 NA NA 9.117621e-02 NA 0 0  
## var 1.014105e+02 NA NA 3.890047e+00 NA 0 0  
## std.dev 1.007028e+01 NA NA 1.972320e+00 NA 0 0  
## coef.var 5.031994e-03 NA NA 3.624922e-01 NA 0 0  
## HPI MHI RMHI poverty population  
## nbr.val 1.800000e+03 1.800000e+03 1.800000e+03 1.800000e+03 1.800000e+03  
## nbr.null 0.000000e+00 0.000000e+00 0.000000e+00 0.000000e+00 0.000000e+00  
## nbr.na 0.000000e+00 0.000000e+00 0.000000e+00 0.000000e+00 0.000000e+00  
## min 7.925000e+01 1.543000e+04 3.491600e+04 2.900000e+00 4.536900e+05  
## max 7.968600e+02 8.634500e+04 8.634500e+04 2.720000e+01 3.955705e+07  
## range 7.176100e+02 7.091500e+04 5.142900e+04 2.430000e+01 3.910336e+07  
## sum 4.682747e+05 7.505899e+07 1.037364e+08 2.312780e+04 1.018931e+10  
## median 2.396650e+02 4.109850e+04 5.678950e+04 1.230000e+01 3.904475e+06  
## mean 2.601526e+02 4.169944e+04 5.763133e+04 1.284878e+01 5.660728e+06  
## SE.mean 2.820962e+00 3.137450e+02 2.245659e+02 8.704734e-02 1.475374e+05  
## CI.mean.0.95 5.532707e+00 6.153429e+02 4.404374e+02 1.707245e-01 2.893626e+05  
## var 1.432409e+04 1.771846e+08 9.077371e+07 1.363903e+01 3.918109e+13  
## std.dev 1.196833e+02 1.331107e+04 9.527524e+03 3.693106e+00 6.259480e+06  
## coef.var 4.600503e-01 3.192147e-01 1.653185e-01 2.874286e-01 1.105773e+00  
## sp500 log\_pop log\_RMHI  
## nbr.val 1800.0000000 1.800000e+03 1.800000e+03  
## nbr.null 50.0000000 0.000000e+00 0.000000e+00  
## nbr.na 0.0000000 0.000000e+00 0.000000e+00  
## min -38.4900000 1.302517e+01 1.046070e+01  
## max 34.1100000 1.749325e+01 1.136611e+01  
## range 72.6000000 4.468085e+00 9.054057e-01  
## sum 17061.0000000 2.711241e+04 1.970673e+04  
## median 11.8950000 1.517763e+01 1.094711e+01  
## mean 9.4783333 1.506245e+01 1.094819e+01  
## SE.mean 0.3699741 2.388216e-02 3.901516e-03  
## CI.mean.0.95 0.7256242 4.683969e-02 7.651979e-03  
## var 246.3855447 1.026644e+00 2.739929e-02  
## std.dev 15.6966730 1.013234e+00 1.655273e-01  
## coef.var 1.6560583 6.726890e-02 1.511915e-02

The dependent variable, value, has a mean of 5.44 and a standard deviation of 1.97. Connecticut and Virginia had the lowest unemployment rate in 2000 at 1.8, while West Virginia was the highest in 1984 at 14.4. The mean HPI was 260.15, with a standard deviation of 119.68. The lowest HPI was found in Wyoming in 1987 at 79.25, while the highest was in Massachusetts in 2018 at 796.86. Poverty rate had a mean of 12.85 and a standard deviation of 3.69. In 1989, Connecticut had the lowest rate of 2.9, and Mississippi had the highest at 27.2 in 1988. The sp500 index mean was 9.48, and the standard deviation was 15.70. In 2008, the lowest index was seen at -38.4, while the highest index was seen in 1995 at 34.11. The logged population mean was 15.06, and the standard deviation was 1.01. Wyoming had the lowest logged population at 13.025 in 1990, California having the highest in 2018 at 17.49. Logged real median household income had a mean of 10.95 and a standard deviation of 0.17. In 2013, Mississippi had the lowest log\_RMHI at 10.46, while Massachusetts had the highest in 2018 at 11.37.

After our data was merged we binned the states into groups based on the mean population of the years in our dataframe (1984-2018). Below are the bins that were created:

Population < 2,000,000 WV, NM, NE, ID, ME, HI, NH, RI, MT, DE, SD, ND, AK, VT, WY

2,000,000 < Population < 5,000,000 MN, AL, LA, CO, SC, KY, OK, OR, CT, IA, MS, KS, AR, UT, NV

5,000,000 < Population < 10,000,000 MI, NJ, GA, NC, VA, MA, IN, WA, TN, MO, WI, MD, AZ

Population > 10,000,000 CA, TX, NY, FL, PA, IL, OH

Each member of our group performed some basic analysis on the states within one of the bins. After performing our initial analysis we chose to focus on 2 states within each bin. The states we chose to focus on are: TX, IL, MO, WA, OR, CO, WV, ME

Below are some of the highlights along with commentary for the states we have chosen to focus on

IL\_data <- merged\_final[ merged\_final$state == "IL", ]

#Descriptive statistics  
stat.desc(IL\_data)

## year period periodName value state quarter annual  
## nbr.val 3.600000e+01 NA NA 36.0000000 NA 36 36  
## nbr.null 0.000000e+00 NA NA 0.0000000 NA 0 0  
## nbr.na 0.000000e+00 NA NA 0.0000000 NA 0 0  
## min 1.984000e+03 NA NA 4.1000000 NA 4 1  
## max 2.018000e+03 NA NA 11.0000000 NA 4 1  
## range 3.400000e+01 NA NA 6.9000000 NA 0 0  
## sum 7.204500e+04 NA NA 231.7000000 NA 144 36  
## median 2.001500e+03 NA NA 6.1000000 NA 4 1  
## mean 2.001250e+03 NA NA 6.4361111 NA 4 1  
## SE.mean 1.701715e+00 NA NA 0.3023493 NA 0 0  
## CI.mean.0.95 3.454665e+00 NA NA 0.6138018 NA 0 0  
## var 1.042500e+02 NA NA 3.2909444 NA 0 0  
## std.dev 1.021029e+01 NA NA 1.8140960 NA 0 0  
## coef.var 5.101956e-03 NA NA 0.2818621 NA 0 0  
## HPI MHI RMHI poverty population  
## nbr.val 36.000000 3.600000e+01 3.600000e+01 36.0000000 3.600000e+01  
## nbr.null 0.000000 0.000000e+00 0.000000e+00 0.0000000 0.000000e+00  
## nbr.na 0.000000 0.000000e+00 0.000000e+00 0.0000000 0.000000e+00  
## min 112.800000 2.375200e+04 5.482800e+04 9.9000000 1.138726e+07  
## max 371.210000 7.014500e+04 7.014500e+04 15.6000000 1.289827e+07  
## range 258.410000 4.639300e+04 1.531700e+04 5.7000000 1.511012e+06  
## sum 9218.510000 1.587843e+06 2.204043e+06 451.4000000 4.422845e+08  
## median 276.190000 4.607050e+04 6.109050e+04 12.6000000 1.250700e+07  
## mean 256.069722 4.410675e+04 6.122342e+04 12.5388889 1.228568e+07  
## SE.mean 13.176322 1.969590e+03 6.769020e+02 0.2635917 9.375302e+04  
## CI.mean.0.95 26.749355 3.998480e+03 1.374184e+03 0.5351197 1.903288e+05  
## var 6250.156368 1.396542e+08 1.649507e+07 2.5013016 3.164266e+11  
## std.dev 79.057930 1.181754e+04 4.061412e+03 1.5815504 5.625181e+05  
## coef.var 0.308736 2.679304e-01 6.633756e-02 0.1261316 4.578648e-02  
## sp500 log\_pop log\_RMHI  
## nbr.val 36.000000 3.600000e+01 3.600000e+01  
## nbr.null 1.000000 0.000000e+00 0.000000e+00  
## nbr.na 0.000000 0.000000e+00 0.000000e+00  
## min -38.490000 1.624801e+01 1.091196e+01  
## max 34.110000 1.637260e+01 1.115832e+01  
## range 72.600000 1.245982e-01 2.463635e-01  
## sum 341.220000 5.876247e+02 3.967264e+02  
## median 11.895000 1.634180e+01 1.102010e+01  
## mean 9.478333 1.632291e+01 1.102018e+01  
## SE.mean 2.652485 7.733796e-03 1.093321e-02  
## CI.mean.0.95 5.384831 1.570044e-02 2.219560e-02  
## var 253.284340 2.153217e-03 4.303264e-03  
## std.dev 15.914909 4.640277e-02 6.559927e-02  
## coef.var 1.679083 2.842801e-03 5.952651e-03

The dependent variable, unemployment, has a mean of 6.44 and a standard deviation of 1.81. The lowest unemployment rate was in 1998 and 1999 at 4.1, while the highest was in 2009 at 11.00. The mean HPI was 256.07, with a standard deviation of 79.06. The lowest HPI was in 1984 at 112.80, while the highest was in 2018 at 371.21. Poverty rate had a mean of 12.54 and a standard deviation of 1.58. In 1999, lowest rate was seen of 9.9, and the highest was 15.6 in 1992. The sp500 index mean was 9.48, and the standard deviation was 15.70. In 2008, the lowest index was seen at -38.4, while the highest index was seen in 1995 at 34.11. The logged population mean was 16.32, and the standard deviation was 0.046. The lowest logged population at 16.248 in 1986, and the highest in 2013 at 16.373. Logged real median household income had a mean of 11.02 and a standard deviation of 0.066. In 1984, Illinois had the lowest log\_RMHI at 10.91, while had the highest in 2018 at 11.16.

#pairwise correlations  
sapply(IL\_data, class)

## year period periodName value state quarter   
## "numeric" "character" "character" "numeric" "character" "numeric"   
## annual HPI MHI RMHI poverty population   
## "numeric" "numeric" "numeric" "numeric" "numeric" "numeric"   
## sp500 log\_pop log\_RMHI   
## "numeric" "numeric" "numeric"

sapply(IL\_data, is.factor)

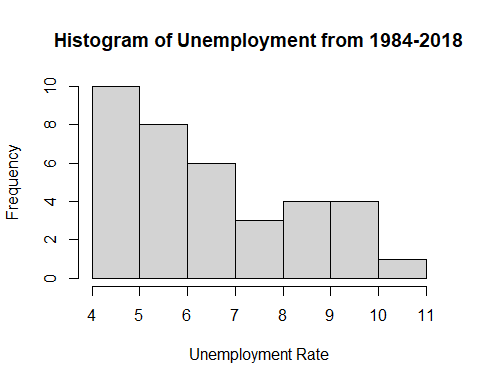
## year period periodName value state quarter annual   
## FALSE FALSE FALSE FALSE FALSE FALSE FALSE   
## HPI MHI RMHI poverty population sp500 log\_pop   
## FALSE FALSE FALSE FALSE FALSE FALSE FALSE   
## log\_RMHI   
## FALSE

cor(IL\_data[sapply(IL\_data, function(x) !is.character(x))])

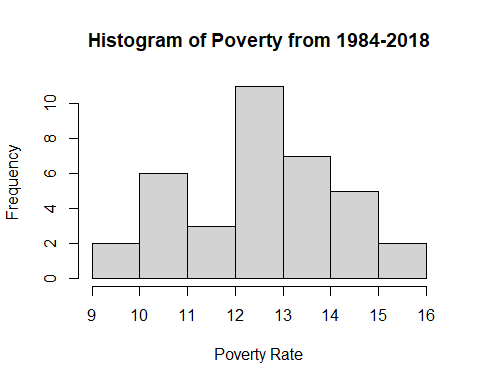
## Warning in cor(IL\_data[sapply(IL\_data, function(x) !is.character(x))]): the  
## standard deviation is zero

## year value quarter annual HPI MHI  
## year 1.000000000 0.006748569 NA NA 0.91598238 0.9754798  
## value 0.006748569 1.000000000 NA NA -0.08115831 -0.1343498  
## quarter NA NA 1 NA NA NA  
## annual NA NA NA 1 NA NA  
## HPI 0.915982382 -0.081158315 NA NA 1.00000000 0.9020434  
## MHI 0.975479844 -0.134349785 NA NA 0.90204345 1.0000000  
## RMHI 0.369670810 -0.659000548 NA NA 0.39861034 0.5518278  
## poverty -0.360502692 0.697578917 NA NA -0.47108192 -0.5052706  
## population 0.955653199 0.020119683 NA NA 0.93700387 0.9194634  
## sp500 -0.138529275 0.031097916 NA NA -0.21926746 -0.1520508  
## log\_pop 0.953215398 0.011231165 NA NA 0.93704886 0.9183431  
## log\_RMHI 0.373224106 -0.660277555 NA NA 0.40824269 0.5536662  
## RMHI poverty population sp500 log\_pop  
## year 0.36967081 -0.36050269 0.95565320 -0.13852928 0.95321540  
## value -0.65900055 0.69757892 0.02011968 0.03109792 0.01123117  
## quarter NA NA NA NA NA  
## annual NA NA NA NA NA  
## HPI 0.39861034 -0.47108192 0.93700387 -0.21926746 0.93704886  
## MHI 0.55182782 -0.50527064 0.91946339 -0.15205085 0.91834308  
## RMHI 1.00000000 -0.86935532 0.36060017 -0.04610779 0.36698125  
## poverty -0.86935532 1.00000000 -0.40348032 0.08946888 -0.41069905  
## population 0.36060017 -0.40348032 1.00000000 -0.15320230 0.99988750  
## sp500 -0.04610779 0.08946888 -0.15320230 1.00000000 -0.15281783  
## log\_pop 0.36698125 -0.41069905 0.99988750 -0.15281783 1.00000000  
## log\_RMHI 0.99934815 -0.87399522 0.36621331 -0.04538721 0.37261639  
## log\_RMHI  
## year 0.37322411  
## value -0.66027755  
## quarter NA  
## annual NA  
## HPI 0.40824269  
## MHI 0.55366622  
## RMHI 0.99934815  
## poverty -0.87399522  
## population 0.36621331  
## sp500 -0.04538721  
## log\_pop 0.37261639  
## log\_RMHI 1.00000000

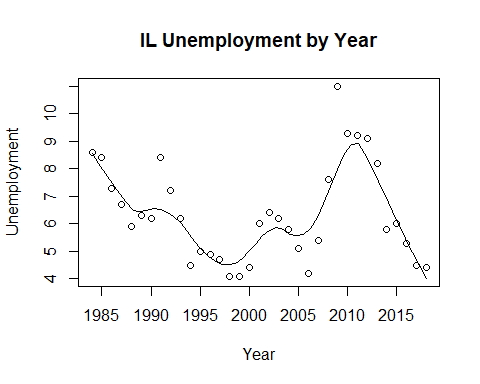
hist(IL\_data$value,  
 main="Histogram of Unemployment from 1984-2018",  
 xlim=c(4,11),  
 xlab="Unemployment Rate",  
 breaks = 5)

 Histograms graphically summarize the distribution of a data set. In Illinois, from 1984 to 2018, The annual unemployment rates ranged from 4.0-11.0%. The histogram displays a positively skewed distribution. An unemployment rate that fell within the first two intervals, 4.0-4.9% & 5.0-5.9% occurred 18 out of the 36 years examined.

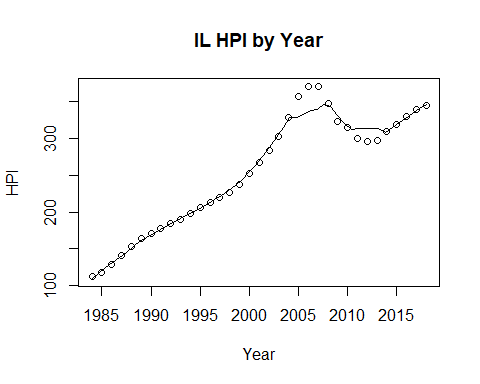
hist(IL\_data$poverty,  
 main="Histogram of Poverty from 1984-2018",  
 xlim=c(9,16),  
 xlab="Poverty Rate",  
 breaks = 5)

 In Illinois, from 1984 to 2018, the poverty levels ranged from 9.0-16.0%. The histogram displays a slightly negatively skewed distribution meaning relatively higher poverty rates occurred more often than lower rates. Between 12.0-12.9% was the most frequently recorded interval during this time period. In 14 of the 36 years examined, the poverty rate in Illinois was between 13-15.9%.

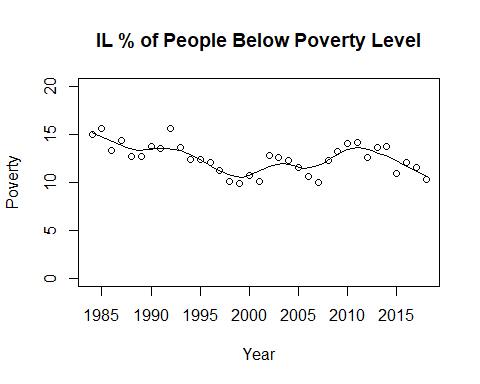
scatter.smooth(IL\_data$year, IL\_data$value, main="IL Unemployment by Year", xlab = "Year", ylab = "Unemployment", span = 2/8)



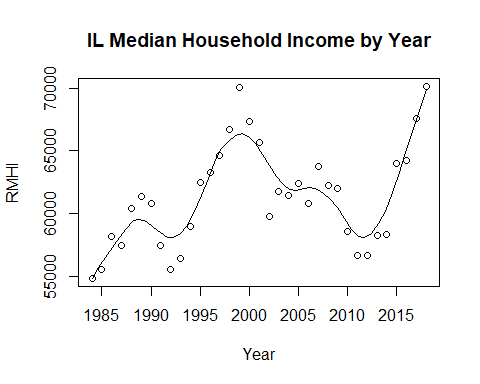
scatter.smooth(IL\_data$year, IL\_data$HPI, main="IL HPI by Year", xlab = "Year", ylab = "HPI", span = 2/8)



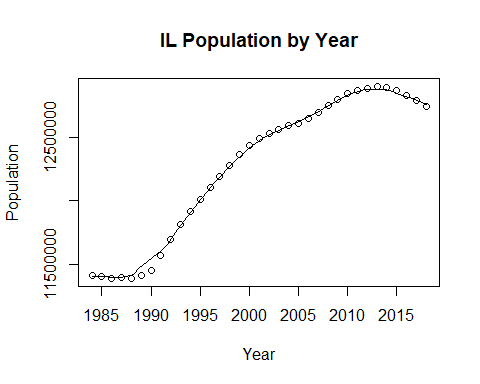
scatter.smooth(x=IL\_data$year, y=IL\_data$poverty, main="IL % of People Below Poverty Level", xlab = "Year", ylab = "Poverty", ylim = c(0,20), span = 2/8)



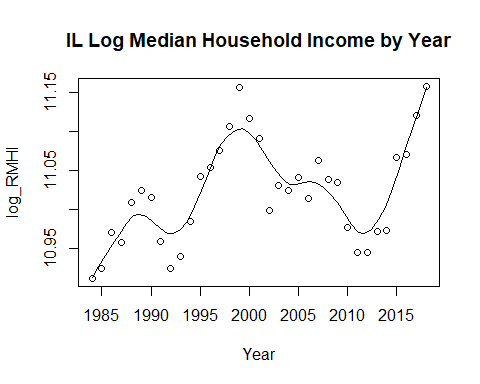
IL\_data$RMHI <- as.integer(IL\_data$RMHI)  
scatter.smooth(x=IL\_data$year, y=IL\_data$RMHI, main="IL Median Household Income by Year", xlab = "Year", ylab = "RMHI", span = 2/8 )



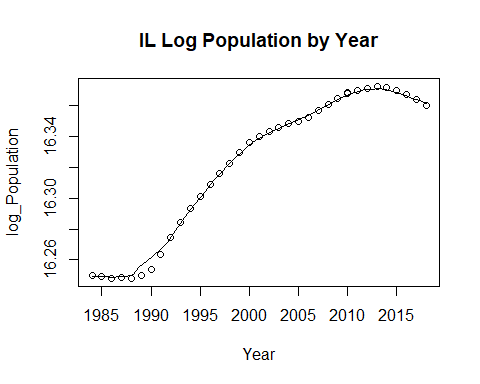
IL\_data$population <- as.integer(IL\_data$population)  
scatter.smooth(x=IL\_data$year, y=IL\_data$population, main="IL Population by Year", xlab = "Year", ylab = "Population", span = 2/8 )



#IL\_data$log\_RMHI <- as.integer(IL\_data$log\_RMHI)  
scatter.smooth(x=IL\_data$year, y=IL\_data$log\_RMHI, main="IL Log Median Household Income by Year", xlab = "Year", ylab = "log\_RMHI", span = 2/8 )



#IL\_data$log\_pop <- as.integer(IL\_data$log\_pop)  
scatter.smooth(x=IL\_data$year, y=IL\_data$log\_pop, main="IL Log Population by Year", xlab = "Year", ylab = "log\_Population", span = 2/8)



cor(IL\_data$RMHI, IL\_data$value)

## [1] -0.6590005

cor(IL\_data$HPI, IL\_data$value)

## [1] -0.08115831

cor(IL\_data$poverty, IL\_data$value)

## [1] 0.6975789

cor(IL\_data$population, IL\_data$value)

## [1] 0.02011968

cor(IL\_data$sp500, IL\_data$value)

## [1] 0.03109792

cor(IL\_data$log\_pop, IL\_data$value)

## [1] 0.01123117

cor(IL\_data$log\_RMHI, IL\_data$value)

## [1] -0.6602776

#run base correlations between categories and dependent variable

A correlation function is used to illustrate the relationship between our independent variable and our dependent variable. The closer the correlation is to 1 or -1 the stronger the correlation between the two variables. For a correlation of 1, it signifies that there is a perfect positive correlation. For a -1 correlation there is a perfect inverse relationship between the two variables.

The correlation between the RMHI and our dependent variable, unemployment, in Illinois is 0.03109792. The correlation shows that there is a weak positive correlation between median household income

The correlation between HPI (Housing Price Index) and our dependent variable, unemployment, in Illinois is -0.08115831. The correlation shows that there is a weak negative relationship between the housing price index and unemployment.

The correlation between poverty, the percentage of the population under the poverty line, and unemployment in Illinois is 0.6975789. This correlation shows a much stronger positive relationship between the two variables. As unemployment increases there is a much stronger possibility that the percentage of the population under the poverty line increasing as well.

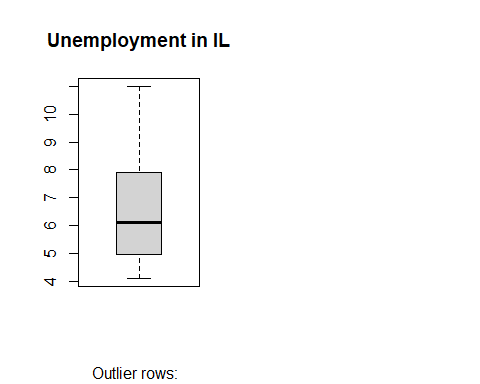
The correlation between the population and unemployment in Illinois is 0.02011968. The correlation between population and unemployment is a weak positive correlation. The correlation is extremely close to 0, which signified that there is no direct correlation between the two variables.

The correlation between the s&p 500 and unemployment in Illinois is 0.03109792. The correlation between the S&P 500 and unemployment is a weak positive correlation. With the correlation being extremely close to 0 it signifies that there is not much of a correlation between the two variables.

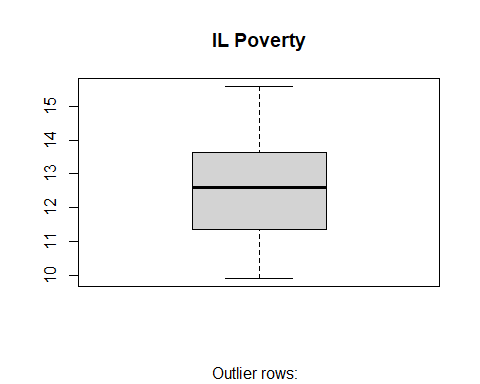
The correlation between the log of the population and unemployment in Illinois is 0.01123117. The correlation between the log of the population and unemployment is extremely close to 0, which signifies there is almost no correlation between the two variables.

The correlation between the log of median household income and unemployment in Illinois is -0.6602776. The correlation between the log of median household income and unemployment is a strong negative correlation. A strong negative correlation suggests that the two variables move in inverse directions of each other. In this case, -0.66 suggests that there is enough evidence to suggest that in most cases the two variables move in inverse directions of each other due to the proximity of -0.66 to -1 which is a perfect negative relationship.

par(mfrow=c(1, 2)) # divide graph area in 2 columns  
boxplot(IL\_data$value, main="Unemployment in IL", sub=paste("Outlier rows: ", boxplot.stats(IL\_data$value)$out)) # box plot for 'Unemployment'



boxplot(IL\_data$poverty, main="IL Poverty", sub=paste("Outlier rows: ", boxplot.stats(IL\_data$poverty)$out)) # box plot for 'Poverty'



#create box plot for poverty level and unemployment level

#run multiple linear model for data  
IL\_reg1 <- lm(value ~ poverty + RMHI + HPI + population + sp500, data = IL\_data)  
summary(IL\_reg1)

##   
## Call:  
## lm(formula = value ~ poverty + RMHI + HPI + population + sp500,   
## data = IL\_data)  
##   
## Residuals:  
## Min 1Q Median 3Q Max   
## -2.5866 -0.7214 -0.0935 0.7777 3.6103   
##   
## Coefficients:  
## Estimate Std. Error t value Pr(>|t|)   
## (Intercept) -1.676e+01 1.379e+01 -1.215 0.2337   
## poverty 6.896e-01 2.776e-01 2.484 0.0188 \*  
## RMHI -1.113e-04 1.033e-04 -1.077 0.2901   
## HPI -5.417e-03 8.027e-03 -0.675 0.5050   
## population 1.851e-06 1.075e-06 1.723 0.0952 .  
## sp500 2.284e-04 1.343e-02 0.017 0.9865   
## ---  
## Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1  
##   
## Residual standard error: 1.217 on 30 degrees of freedom  
## Multiple R-squared: 0.6143, Adjusted R-squared: 0.5501   
## F-statistic: 9.558 on 5 and 30 DF, p-value: 1.561e-05

anova(IL\_reg1)

## Analysis of Variance Table  
##   
## Response: value  
## Df Sum Sq Mean Sq F value Pr(>F)   
## poverty 1 56.050 56.050 37.8541 9.109e-07 \*\*\*  
## RMHI 1 1.303 1.303 0.8798 0.35574   
## HPI 1 8.899 8.899 6.0102 0.02026 \*   
## population 1 4.510 4.510 3.0461 0.09117 .   
## sp500 1 0.000 0.000 0.0003 0.98654   
## Residuals 30 44.420 1.481   
## ---  
## Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1

IL\_reg2 <- lm(value ~ HPI + poverty + sp500 + log\_pop + log\_RMHI , data = IL\_data)  
summary(IL\_reg2)

##   
## Call:  
## lm(formula = value ~ HPI + poverty + sp500 + log\_pop + log\_RMHI,   
## data = IL\_data)  
##   
## Residuals:  
## Min 1Q Median 3Q Max   
## -2.5575 -0.7142 -0.0868 0.7908 3.6217   
##   
## Coefficients:  
## Estimate Std. Error t value Pr(>|t|)   
## (Intercept) -2.701e+02 2.170e+02 -1.245 0.2228   
## HPI -4.554e-03 7.999e-03 -0.569 0.5733   
## poverty 6.973e-01 2.810e-01 2.481 0.0189 \*  
## sp500 5.088e-04 1.349e-02 0.038 0.9702   
## log\_pop 2.113e+01 1.303e+01 1.621 0.1154   
## log\_RMHI -6.889e+00 6.524e+00 -1.056 0.2994   
## ---  
## Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1  
##   
## Residual standard error: 1.222 on 30 degrees of freedom  
## Multiple R-squared: 0.6109, Adjusted R-squared: 0.546   
## F-statistic: 9.418 on 5 and 30 DF, p-value: 1.773e-05

anova(IL\_reg2)

## Analysis of Variance Table  
##   
## Response: value  
## Df Sum Sq Mean Sq F value Pr(>F)   
## HPI 1 0.759 0.759 0.5078 0.4816   
## poverty 1 64.356 64.356 43.0732 2.924e-07 \*\*\*  
## sp500 1 0.076 0.076 0.0507 0.8234   
## log\_pop 1 3.503 3.503 2.3446 0.1362   
## log\_RMHI 1 1.666 1.666 1.1149 0.2994   
## Residuals 30 44.823 1.494   
## ---  
## Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1

IL\_linearModelSignificant <- lm(value ~ poverty + log\_pop, data = IL\_data)  
summary(IL\_linearModelSignificant)

##   
## Call:  
## lm(formula = value ~ poverty + log\_pop, data = IL\_data)  
##   
## Residuals:  
## Min 1Q Median 3Q Max   
## -2.4468 -0.7344 -0.1555 0.7208 3.3382   
##   
## Coefficients:  
## Estimate Std. Error t value Pr(>|t|)   
## (Intercept) -234.2508 78.4390 -2.986 0.00529 \*\*   
## poverty 0.9689 0.1397 6.937 6.3e-08 \*\*\*  
## log\_pop 14.0011 4.7604 2.941 0.00594 \*\*   
## ---  
## Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1  
##   
## Residual standard error: 1.192 on 33 degrees of freedom  
## Multiple R-squared: 0.5932, Adjusted R-squared: 0.5686   
## F-statistic: 24.06 on 2 and 33 DF, p-value: 3.581e-07

TX\_data <- merged\_final[ merged\_final$state == "TX", ]

#Descriptive statistics  
stat.desc(TX\_data)

## year period periodName value state quarter annual  
## nbr.val 3.600000e+01 NA NA 36.0000000 NA 36 36  
## nbr.null 0.000000e+00 NA NA 0.0000000 NA 0 0  
## nbr.na 0.000000e+00 NA NA 0.0000000 NA 0 0  
## min 1.984000e+03 NA NA 3.5000000 NA 4 1  
## max 2.018000e+03 NA NA 8.5000000 NA 4 1  
## range 3.400000e+01 NA NA 5.0000000 NA 0 0  
## sum 7.204500e+04 NA NA 205.1000000 NA 144 36  
## median 2.001500e+03 NA NA 5.6500000 NA 4 1  
## mean 2.001250e+03 NA NA 5.6972222 NA 4 1  
## SE.mean 1.701715e+00 NA NA 0.2254268 NA 0 0  
## CI.mean.0.95 3.454665e+00 NA NA 0.4576407 NA 0 0  
## var 1.042500e+02 NA NA 1.8294206 NA 0 0  
## std.dev 1.021029e+01 NA NA 1.3525608 NA 0 0  
## coef.var 5.101956e-03 NA NA 0.2374071 NA 0 0  
## HPI MHI RMHI poverty population  
## nbr.val 36.0000000 3.600000e+01 3.600000e+01 36.00000000 3.600000e+01  
## nbr.null 0.0000000 0.000000e+00 0.000000e+00 0.00000000 0.000000e+00  
## nbr.na 0.0000000 0.000000e+00 0.000000e+00 0.00000000 0.000000e+00  
## min 113.7700000 2.302400e+04 4.917700e+04 13.20000000 1.600709e+07  
## max 334.9200000 6.009200e+04 6.155700e+04 19.10000000 2.870185e+07  
## range 221.1500000 3.706800e+04 1.238000e+04 5.90000000 1.269476e+07  
## sum 6687.8500000 1.424867e+06 1.969117e+06 592.50000000 7.822428e+08  
## median 180.4800000 3.971000e+04 5.455400e+04 16.55000000 2.150497e+07  
## mean 185.7736111 3.957964e+04 5.469769e+04 16.45833333 2.172897e+07  
## SE.mean 10.2226857 1.881071e+03 5.124132e+02 0.22801333 6.691423e+05  
## CI.mean.0.95 20.7531553 3.818777e+03 1.040254e+03 0.46289167 1.358431e+06  
## var 3762.1189266 1.273834e+08 9.452421e+06 1.87164286 1.611905e+13  
## std.dev 61.3361144 1.128642e+04 3.074479e+03 1.36807999 4.014854e+06  
## coef.var 0.3301659 2.851573e-01 5.620857e-02 0.08312385 1.847697e-01  
## sp500 log\_pop log\_RMHI  
## nbr.val 36.000000 36.00000000 3.600000e+01  
## nbr.null 1.000000 0.00000000 0.000000e+00  
## nbr.na 0.000000 0.00000000 0.000000e+00  
## min -38.490000 16.58854206 1.080318e+01  
## max 34.110000 17.17247196 1.102772e+01  
## range 72.600000 0.58392991 2.245375e-01  
## sum 341.220000 607.58704791 3.926899e+02  
## median 11.895000 16.88375765 1.090695e+01  
## mean 9.478333 16.87741800 1.090805e+01  
## SE.mean 2.652485 0.03100954 9.311408e-03  
## CI.mean.0.95 5.384831 0.06295270 1.890316e-02  
## var 253.284340 0.03461729 3.121284e-03  
## std.dev 15.914909 0.18605721 5.586845e-02  
## coef.var 1.679083 0.01102403 5.121761e-03

The dependent variable, unemployment, has a mean of 5.70 and a standard deviation of 1.35. The lowest unemployment rate was in 2000 at 3.5, while the highest was in 1986 at 8.5. The mean HPI was 185.77, with a standard deviation of 61.34. The lowest HPI was in 1988 at 113.77, while the highest was in 2018 at 334.92. Poverty rate had a mean of 16.46 and a standard deviation of 1.37. In 2017, lowest rate was seen of 13.2 and the highest was 19.1 in 1994. The sp500 index mean was 9.48, and the standard deviation was 15.70. In 2008, the lowest index was seen at -38.4, while the highest index was seen in 1995 at 34.11. The logged population mean was 16.88, and the standard deviation was 0.17. The lowest logged population at 16.59 in 1984, and the highest in 2018 at 17.17. Logged real median household income had a mean of 10.91 and a standard deviation of 0.056. In 1992, Texas had the lowest log\_RMHI at 10.80, while had the highest in 2017 at 11.03.

#pairwise correlations  
sapply(TX\_data, class)

## year period periodName value state quarter   
## "numeric" "character" "character" "numeric" "character" "numeric"   
## annual HPI MHI RMHI poverty population   
## "numeric" "numeric" "numeric" "numeric" "numeric" "numeric"   
## sp500 log\_pop log\_RMHI   
## "numeric" "numeric" "numeric"

sapply(TX\_data, is.factor)

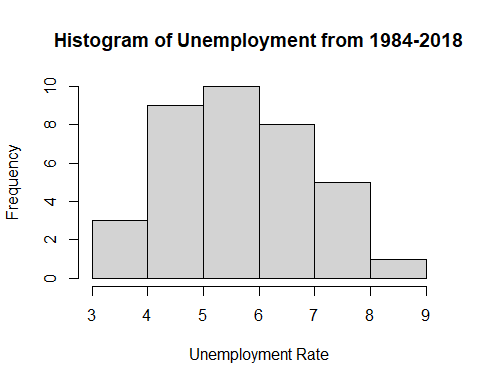
## year period periodName value state quarter annual   
## FALSE FALSE FALSE FALSE FALSE FALSE FALSE   
## HPI MHI RMHI poverty population sp500 log\_pop   
## FALSE FALSE FALSE FALSE FALSE FALSE FALSE   
## log\_RMHI   
## FALSE

cor(TX\_data[sapply(TX\_data, function(x) !is.character(x))])

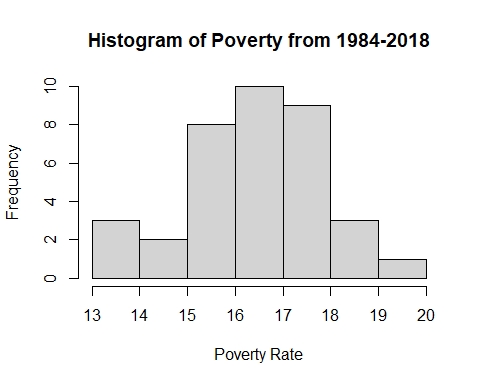
## Warning in cor(TX\_data[sapply(TX\_data, function(x) !is.character(x))]): the  
## standard deviation is zero

## year value quarter annual HPI MHI  
## year 1.0000000 -0.40586425 NA NA 0.9490575 0.9917495  
## value -0.4058642 1.00000000 NA NA -0.4234252 -0.4429500  
## quarter NA NA 1 NA NA NA  
## annual NA NA NA 1 NA NA  
## HPI 0.9490575 -0.42342522 NA NA 1.0000000 0.9685900  
## MHI 0.9917495 -0.44295000 NA NA 0.9685900 1.0000000  
## RMHI 0.7549419 -0.56630739 NA NA 0.8086748 0.8217689  
## poverty -0.3839764 0.67190981 NA NA -0.5300500 -0.4713178  
## population 0.9951277 -0.38612661 NA NA 0.9679242 0.9948080  
## sp500 -0.1385293 0.09162901 NA NA -0.1740205 -0.1513208  
## log\_pop 0.9983833 -0.39411107 NA NA 0.9529186 0.9921431  
## log\_RMHI 0.7550869 -0.56262851 NA NA 0.8048340 0.8205671  
## RMHI poverty population sp500 log\_pop log\_RMHI  
## year 0.7549419 -0.3839764 0.9951277 -0.13852928 0.9983833 0.7550869  
## value -0.5663074 0.6719098 -0.3861266 0.09162901 -0.3941111 -0.5626285  
## quarter NA NA NA NA NA NA  
## annual NA NA NA NA NA NA  
## HPI 0.8086748 -0.5300500 0.9679242 -0.17402046 0.9529186 0.8048340  
## MHI 0.8217689 -0.4713178 0.9948080 -0.15132081 0.9921431 0.8205671  
## RMHI 1.0000000 -0.7661225 0.7761129 -0.15576943 0.7672077 0.9993853  
## poverty -0.7661225 1.0000000 -0.4049827 0.21354689 -0.3874947 -0.7579950  
## population 0.7761129 -0.4049827 1.0000000 -0.13537154 0.9971638 0.7760053  
## sp500 -0.1557694 0.2135469 -0.1353715 1.00000000 -0.1408642 -0.1571481  
## log\_pop 0.7672077 -0.3874947 0.9971638 -0.14086421 1.0000000 0.7683718  
## log\_RMHI 0.9993853 -0.7579950 0.7760053 -0.15714808 0.7683718 1.0000000

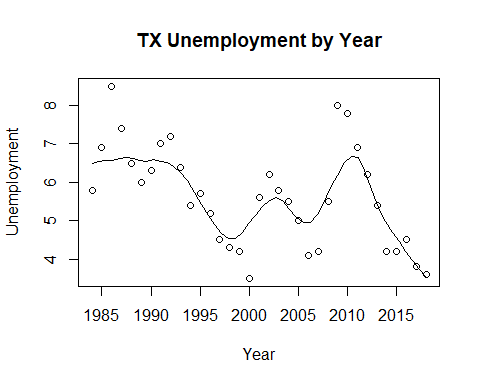
hist(TX\_data$value,  
 main="Histogram of Unemployment from 1984-2018",  
 xlim=c(3,9),  
 xlab="Unemployment Rate",  
 breaks = 5)

 In Texas, from 1984 to 2018, the annual unemployment rates ranged from 3.0-8.9%. The most common interval, or range of annual unemployment rates was between 5.0-5.9% which occurred 10 times in the 36 year period. Texas only experienced one year with an unemployment rate of 8.0% or higher.

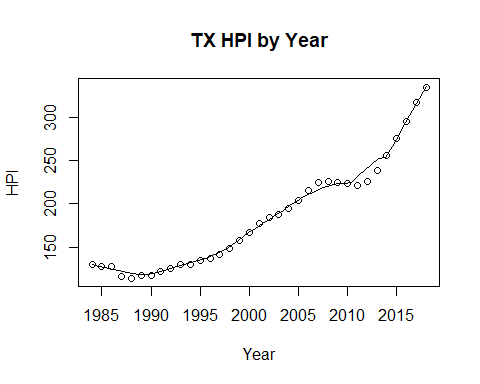
hist(TX\_data$poverty,  
 main="Histogram of Poverty from 1984-2018",  
 xlim=c(13,20),  
 xlab="Poverty Rate",  
 breaks = 5)

 In Texas, from 1984 to 2018, the annual poverty rates ranged from 13.0-20.0%. The histogram is nearly symmetrical meaning most of the values occurred in the near middle of the range. 27 of the 36 years were characterized by poverty rates ranging from 15.0-17.9%. Only 5 years experienced lower rates and only 4 experienced higher rates of poverty.

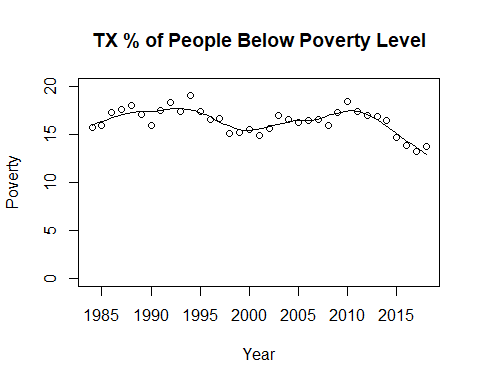
scatter.smooth(TX\_data$year, TX\_data$value, main="TX Unemployment by Year", xlab = "Year", ylab = "Unemployment", span = 2/8)



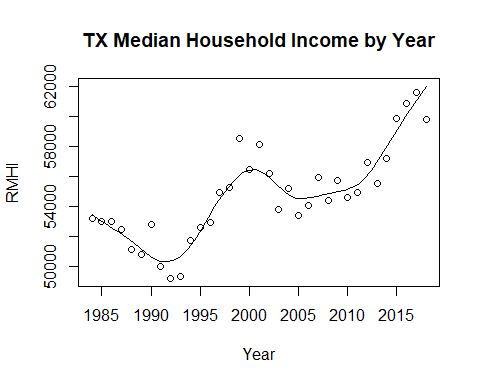
scatter.smooth(TX\_data$year, TX\_data$HPI, main="TX HPI by Year", xlab = "Year", ylab = "HPI", span = 2/8)



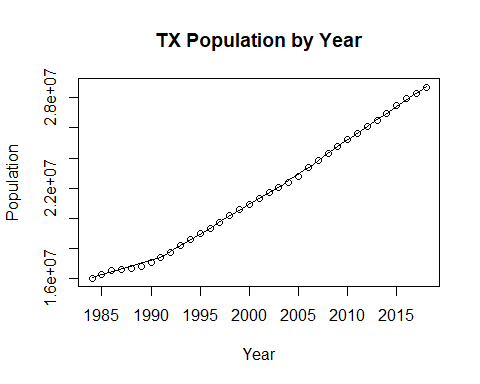
scatter.smooth(x=TX\_data$year, y=TX\_data$poverty, main="TX % of People Below Poverty Level", xlab = "Year", ylab = "Poverty", ylim = c(0,20), span = 2/8)



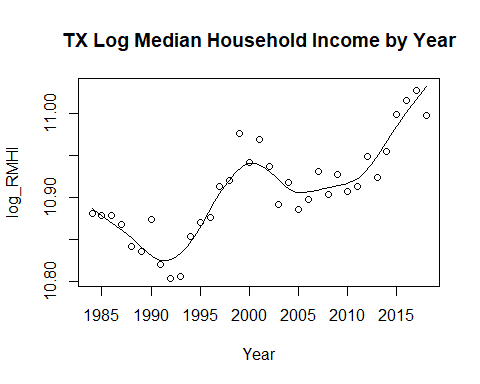
TX\_data$RMHI <- as.integer(TX\_data$RMHI)  
scatter.smooth(x=TX\_data$year, y=TX\_data$RMHI, main="TX Median Household Income by Year", xlab = "Year", ylab = "RMHI", span = 2/8 )



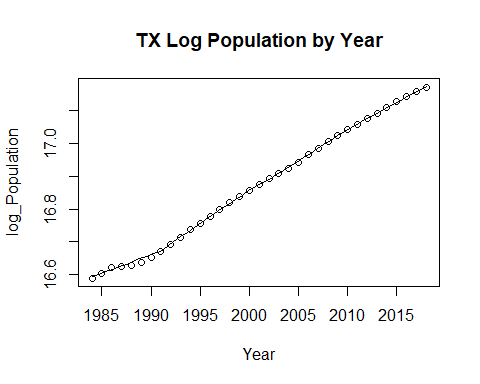
TX\_data$population <- as.integer(TX\_data$population)  
scatter.smooth(x=TX\_data$year, y=TX\_data$population, main="TX Population by Year", xlab = "Year", ylab = "Population", span = 2/8 )



#TX\_data$log\_RMHI <- as.integer(TX\_data$log\_RMHI)  
scatter.smooth(x=TX\_data$year, y=TX\_data$log\_RMHI, main="TX Log Median Household Income by Year", xlab = "Year", ylab = "log\_RMHI", span = 2/8 )



#TX\_data$log\_pop <- as.integer(TX\_data$log\_pop)  
scatter.smooth(x=TX\_data$year, y=TX\_data$log\_pop, main="TX Log Population by Year", xlab = "Year", ylab = "log\_Population", span = 2/8)



cor(TX\_data$RMHI, TX\_data$value)

## [1] -0.5663074

cor(TX\_data$HPI, TX\_data$value)

## [1] -0.4234252

cor(TX\_data$poverty, TX\_data$value)

## [1] 0.6719098

cor(TX\_data$population, TX\_data$value)

## [1] -0.3861266

cor(TX\_data$sp500, TX\_data$value)

## [1] 0.09162901

cor(TX\_data$log\_pop, TX\_data$value)

## [1] -0.3941111

cor(TX\_data$log\_RMHI, TX\_data$value)

## [1] -0.5626285

#run base correlations between categories and dependent variable

The correlation between the RMHI and our dependent variable, unemployment, in Texas is -0.5663074. The correlation between median household income and unemployment is a decently strong negative correlation. This is much stronger correlation than was the vase with the previous state of Illinois.

The correlation between the housing price index and unemployment in Texas is -0.4234252 which is a negative correlation. It illustrates a correlation that points towards an inverse relationship between the housing price index and unemployment.

The correlation between poverty, the proportion of the population under the poverty line, and unemployment in Texas is 0.6719098. The correlation between poverty and unemployment shows a decently strong correlation between the two variables. The positive correlation shows that the variables move in the same direction dependent on the value of the variables.

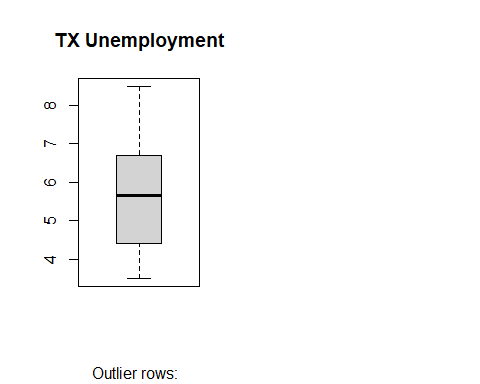
The correlation between population and unemployment in Texas is -0.3861266. The correlation between population and unemployment points to a weak inverse relationship. This means that as population goes up, then the unemployment tends to go down. The correlation is not as close to -1 as it is to 0, which shows that there are instances where there is no direct relationship between the two variables.

The correlation between the S&P 500 and unemployment in Texas is 0.09162901. The correlation between S&P 500 and unemployment is a weak positive correlation. With the correlation being this close to zero it shows that in most cases there is no direct correlation between the two variables.

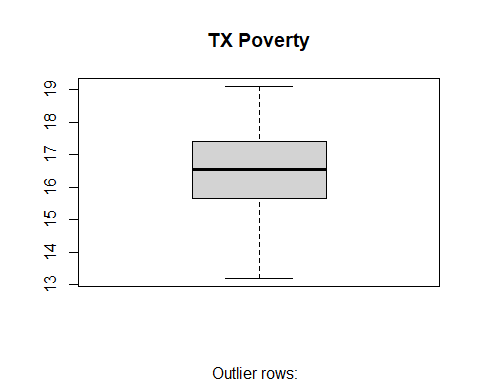
The correlation between the log of the population and unemployment in Texas is -0.3941111. The correlation between the log of population and unemployment shows that there is an inverse relationship between the two variables. THe correlation is not the strongest, but does illustrate that in certain cases the two variables do have an inverse relationship.

The correlation between the log of the median household income and unemployment in Texas is -0.5626285. THe correlation between the log of the median household income and unemployment shows a strong inverse relationship. An inverse relationship suggests that the two variables tend to go in opposite directions of each other. So, for instance, as median household income goes up,then the unemployment tends to go down.

par(mfrow=c(1, 2)) # divide graph area in 2 columns  
boxplot(TX\_data$value, main="TX Unemployment", sub=paste("Outlier rows: ", boxplot.stats(TX\_data$value)$out)) # box plot for 'Unemployment'



boxplot(TX\_data$poverty, main="TX Poverty", sub=paste("Outlier rows: ", boxplot.stats(TX\_data$poverty)$out)) # box plot for 'Poverty'



#create box plot for poverty level and unemployment level

#run multiple linear model for data  
TX\_reg1 <- lm(value ~ poverty + RMHI + HPI + population + sp500, data = TX\_data)  
summary(TX\_reg1)

##   
## Call:  
## lm(formula = value ~ poverty + RMHI + HPI + population + sp500,   
## data = TX\_data)  
##   
## Residuals:  
## Min 1Q Median 3Q Max   
## -2.42022 -0.50730 0.00617 0.69676 1.87937   
##   
## Coefficients:  
## Estimate Std. Error t value Pr(>|t|)   
## (Intercept) -9.230e+00 1.015e+01 -0.909 0.37042   
## poverty 8.773e-01 2.687e-01 3.265 0.00274 \*\*  
## RMHI 8.587e-05 1.481e-04 0.580 0.56640   
## HPI 2.074e-02 1.448e-02 1.433 0.16225   
## population -3.691e-07 2.264e-07 -1.631 0.11340   
## sp500 -4.429e-03 1.126e-02 -0.393 0.69675   
## ---  
## Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1  
##   
## Residual standard error: 1.027 on 30 degrees of freedom  
## Multiple R-squared: 0.5057, Adjusted R-squared: 0.4234   
## F-statistic: 6.14 on 5 and 30 DF, p-value: 0.0004974

anova(TX\_reg1)

## Analysis of Variance Table  
##   
## Response: value  
## Df Sum Sq Mean Sq F value Pr(>F)   
## poverty 1 28.907 28.9070 27.4029 1.2e-05 \*\*\*  
## RMHI 1 0.412 0.4118 0.3904 0.5368   
## HPI 1 0.057 0.0569 0.0540 0.8179   
## population 1 2.844 2.8439 2.6959 0.1111   
## sp500 1 0.163 0.1633 0.1548 0.6968   
## Residuals 30 31.647 1.0549   
## ---  
## Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1

TX\_reg2 <- lm(value ~ HPI + poverty + sp500 + log\_pop + log\_RMHI, data = TX\_data)  
summary(TX\_reg2)

##   
## Call:  
## lm(formula = value ~ HPI + poverty + sp500 + log\_pop + log\_RMHI,   
## data = TX\_data)  
##   
## Residuals:  
## Min 1Q Median 3Q Max   
## -2.44168 -0.44509 0.03532 0.65192 1.92015   
##   
## Coefficients:  
## Estimate Std. Error t value Pr(>|t|)   
## (Intercept) 56.107242 79.166345 0.709 0.48397   
## HPI 0.018999 0.011315 1.679 0.10353   
## poverty 0.897958 0.249964 3.592 0.00115 \*\*  
## sp500 -0.005457 0.011027 -0.495 0.62428   
## log\_pop -7.602177 3.827211 -1.986 0.05620 .   
## log\_RMHI 5.467371 7.723899 0.708 0.48450   
## ---  
## Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1  
##   
## Residual standard error: 1.007 on 30 degrees of freedom  
## Multiple R-squared: 0.5247, Adjusted R-squared: 0.4454   
## F-statistic: 6.622 on 5 and 30 DF, p-value: 0.0002909

anova(TX\_reg2)

## Analysis of Variance Table  
##   
## Response: value  
## Df Sum Sq Mean Sq F value Pr(>F)   
## HPI 1 11.4798 11.4798 11.3152 0.0021154 \*\*   
## poverty 1 17.8303 17.8303 17.5747 0.0002246 \*\*\*  
## sp500 1 0.2234 0.2234 0.2202 0.6422971   
## log\_pop 1 3.5515 3.5515 3.5006 0.0711299 .   
## log\_RMHI 1 0.5083 0.5083 0.5011 0.4845011   
## Residuals 30 30.4364 1.0145   
## ---  
## Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1

TX\_linearModelSignificant <- lm(value ~ poverty + log\_pop, data = TX\_data)  
summary(TX\_linearModelSignificant)

##   
## Call:  
## lm(formula = value ~ poverty + log\_pop, data = TX\_data)  
##   
## Residuals:  
## Min 1Q Median 3Q Max   
## -2.05373 -0.65686 -0.07282 0.71777 2.00284   
##   
## Coefficients:  
## Estimate Std. Error t value Pr(>|t|)   
## (Intercept) 15.0658 17.8113 0.846 0.404   
## poverty 0.6040 0.1356 4.455 9.11e-05 \*\*\*  
## log\_pop -1.1441 0.9970 -1.148 0.259   
## ---  
## Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1  
##   
## Residual standard error: 1.012 on 33 degrees of freedom  
## Multiple R-squared: 0.4725, Adjusted R-squared: 0.4405   
## F-statistic: 14.78 on 2 and 33 DF, p-value: 2.609e-05

MO\_data <- merged\_final[ merged\_final$state == "MO", ]

#Descriptive statistics  
stat.desc(MO\_data)

## year period periodName value state quarter annual  
## nbr.val 3.600000e+01 NA NA 36.0000000 NA 36 36  
## nbr.null 0.000000e+00 NA NA 0.0000000 NA 0 0  
## nbr.na 0.000000e+00 NA NA 0.0000000 NA 0 0  
## min 1.984000e+03 NA NA 2.8000000 NA 4 1  
## max 2.018000e+03 NA NA 9.4000000 NA 4 1  
## range 3.400000e+01 NA NA 6.6000000 NA 0 0  
## sum 7.204500e+04 NA NA 194.5000000 NA 144 36  
## median 2.001500e+03 NA NA 5.2000000 NA 4 1  
## mean 2.001250e+03 NA NA 5.4027778 NA 4 1  
## SE.mean 1.701715e+00 NA NA 0.2712840 NA 0 0  
## CI.mean.0.95 3.454665e+00 NA NA 0.5507358 NA 0 0  
## var 1.042500e+02 NA NA 2.6494206 NA 0 0  
## std.dev 1.021029e+01 NA NA 1.6277041 NA 0 0  
## coef.var 5.101956e-03 NA NA 0.3012717 NA 0 0  
## HPI MHI RMHI poverty population  
## nbr.val 36.0000000 3.600000e+01 3.600000e+01 36.0000000 3.600000e+01  
## nbr.null 0.0000000 0.000000e+00 0.000000e+00 0.0000000 0.000000e+00  
## nbr.na 0.0000000 0.000000e+00 0.000000e+00 0.0000000 0.000000e+00  
## min 118.9400000 2.077500e+04 4.795600e+04 9.2000000 4.975278e+06  
## max 334.6000000 6.172600e+04 6.594300e+04 17.5000000 6.126452e+06  
## range 215.6600000 4.095100e+04 1.798700e+04 8.3000000 1.151174e+06  
## sum 8063.2700000 1.433078e+06 1.978707e+06 461.1000000 2.020229e+08  
## median 235.4550000 4.245650e+04 5.509500e+04 12.7500000 5.657984e+06  
## mean 223.9797222 3.980772e+04 5.496408e+04 12.8083333 5.611748e+06  
## SE.mean 10.9568132 1.909017e+03 8.226538e+02 0.3788113 6.452331e+04  
## CI.mean.0.95 22.2435133 3.875511e+03 1.670076e+03 0.7690279 1.309893e+05  
## var 4321.8631913 1.311965e+08 2.436334e+07 5.1659286 1.498773e+11  
## std.dev 65.7408791 1.145410e+04 4.935923e+03 2.2728679 3.871399e+05  
## coef.var 0.2935126 2.877357e-01 8.980270e-02 0.1774523 6.898739e-02  
## sp500 log\_pop log\_RMHI  
## nbr.val 36.000000 3.600000e+01 3.600000e+01  
## nbr.null 1.000000 0.000000e+00 0.000000e+00  
## nbr.na 0.000000 0.000000e+00 0.000000e+00  
## min -38.490000 1.541999e+01 1.077804e+01  
## max 34.110000 1.562813e+01 1.109655e+01  
## range 72.600000 2.081345e-01 3.185068e-01  
## sum 341.220000 5.593689e+02 3.927797e+02  
## median 11.895000 1.554857e+01 1.091680e+01  
## mean 9.478333 1.553802e+01 1.091055e+01  
## SE.mean 2.652485 1.162907e-02 1.488114e-02  
## CI.mean.0.95 5.384831 2.360826e-02 3.021032e-02  
## var 253.284340 4.868468e-03 7.972142e-03  
## std.dev 15.914909 6.977440e-02 8.928685e-02  
## coef.var 1.679083 4.490558e-03 8.183536e-03

The dependent variable, unemployment, has a mean of 5.40 and a standard deviation of 1.63. The lowest unemployment rate was in 1999 at 2.8, while the highest was in 2009 at 9.4. The mean HPI was 223.98, with a standard deviation of 65.74. The lowest HPI was in 1984 at 118.94, while the highest was in 2018 at 334.60. Poverty rate had a mean of 12.81 and a standard deviation of 2.27. In 2000, lowest rate was seen of 9.2, and the highest was 17.5 in 2013. The sp500 index mean was 9.48, and the standard deviation was 15.70. In 2008, the lowest index was seen at -38.4, while the highest index was seen in 1995 at 34.11. The logged population mean was 15.54, and the standard deviation was 0.070. The lowest logged population at 15.42 in 1984, and the highest in 2018 at 15.63. Logged real median household income had a mean of 10.91 and a standard deviation of 0.090. In 1984, Missouri had the lowest log\_RMHI at 10.78, while had the highest in 2000 at 11.10.

#pairwise correlations  
sapply(MO\_data, class)

## year period periodName value state quarter   
## "numeric" "character" "character" "numeric" "character" "numeric"   
## annual HPI MHI RMHI poverty population   
## "numeric" "numeric" "numeric" "numeric" "numeric" "numeric"   
## sp500 log\_pop log\_RMHI   
## "numeric" "numeric" "numeric"

sapply(MO\_data, is.factor)

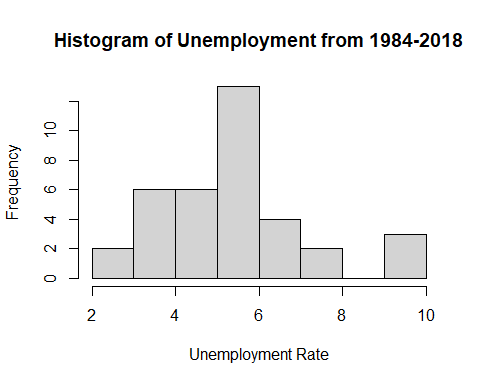
## year period periodName value state quarter annual   
## FALSE FALSE FALSE FALSE FALSE FALSE FALSE   
## HPI MHI RMHI poverty population sp500 log\_pop   
## FALSE FALSE FALSE FALSE FALSE FALSE FALSE   
## log\_RMHI   
## FALSE

cor(MO\_data[sapply(MO\_data, function(x) !is.character(x))])

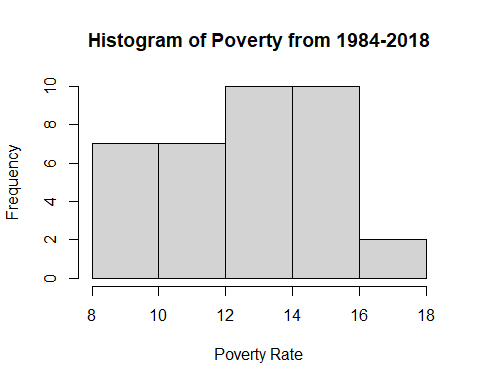
## Warning in cor(MO\_data[sapply(MO\_data, function(x) !is.character(x))]): the  
## standard deviation is zero

## year value quarter annual HPI MHI  
## year 1.00000000 0.03932598 NA NA 0.96247567 0.9726852  
## value 0.03932598 1.00000000 NA NA 0.05064891 -0.1003401  
## quarter NA NA 1 NA NA NA  
## annual NA NA NA 1 NA NA  
## HPI 0.96247567 0.05064891 NA NA 1.00000000 0.9414550  
## MHI 0.97268516 -0.10034014 NA NA 0.94145501 1.0000000  
## RMHI 0.53539838 -0.48320728 NA NA 0.54023310 0.7065757  
## poverty -0.08627458 0.60392697 NA NA -0.14229157 -0.2678414  
## population 0.99237309 0.08687034 NA NA 0.97055778 0.9645918  
## sp500 -0.13852928 -0.09530716 NA NA -0.21070547 -0.1582891  
## log\_pop 0.99028187 0.08057646 NA NA 0.97005431 0.9644376  
## log\_RMHI 0.55197300 -0.47258868 NA NA 0.55920640 0.7191751  
## RMHI poverty population sp500 log\_pop  
## year 0.53539838 -0.08627458 0.99237309 -0.13852928 0.99028187  
## value -0.48320728 0.60392697 0.08687034 -0.09530716 0.08057646  
## quarter NA NA NA NA NA  
## annual NA NA NA NA NA  
## HPI 0.54023310 -0.14229157 0.97055778 -0.21070547 0.97005431  
## MHI 0.70657575 -0.26784136 0.96459181 -0.15828913 0.96443763  
## RMHI 1.00000000 -0.74240102 0.54710579 -0.09670204 0.55761992  
## poverty -0.74240102 1.00000000 -0.09707523 0.07310805 -0.10826802  
## population 0.54710579 -0.09707523 1.00000000 -0.14829752 0.99966603  
## sp500 -0.09670204 0.07310805 -0.14829752 1.00000000 -0.14884087  
## log\_pop 0.55761992 -0.10826802 0.99966603 -0.14884087 1.00000000  
## log\_RMHI 0.99905007 -0.74145337 0.56470044 -0.09169904 0.57538155  
## log\_RMHI  
## year 0.55197300  
## value -0.47258868  
## quarter NA  
## annual NA  
## HPI 0.55920640  
## MHI 0.71917505  
## RMHI 0.99905007  
## poverty -0.74145337  
## population 0.56470044  
## sp500 -0.09169904  
## log\_pop 0.57538155  
## log\_RMHI 1.00000000

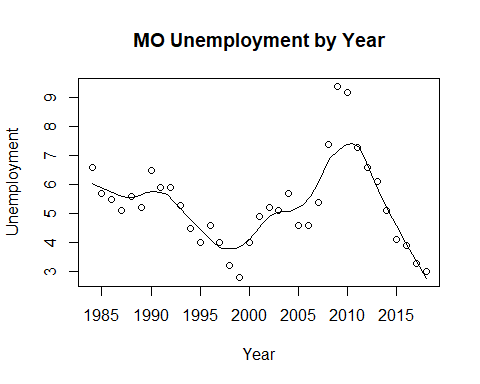
hist(MO\_data$value,  
 main="Histogram of Unemployment from 1984-2018",  
 xlab="Unemployment Rate",  
 breaks = 5)

 In Missouri, from 1984 to 2018, the annual unemployment rates ranged from 2.0-9.9% The most frequent range of unemployment rates was between 5.0-5.9% which occurred 13 out of the 36 years. The next most frequent ranges were between 3.0-3.9% & 4.0-4.9% which both only appeared 6 times each.

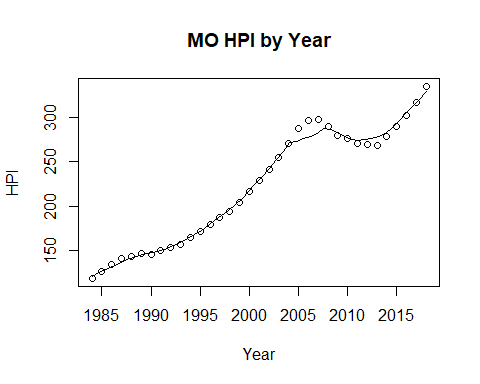
hist(MO\_data$poverty,  
 main="Histogram of Poverty from 1984-2018",  
 xlab="Poverty Rate",  
 breaks = 5)

 In Missouri, from 1984-2014, the annual poverty rates ranged from 8.0-17.9%. 20 of the 36 years were characterized by poverty rates between 12.0-15.9%. Only two years displayed higher poverty rates.

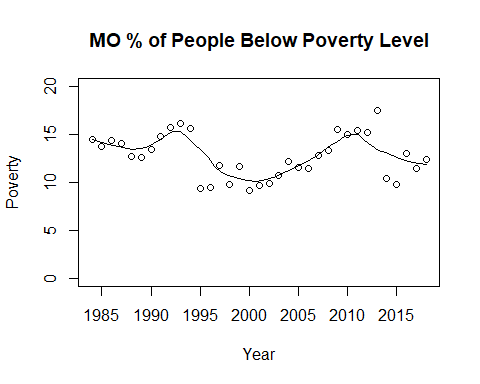
scatter.smooth(MO\_data$year, MO\_data$value, main="MO Unemployment by Year", xlab = "Year", ylab = "Unemployment", span = 2/8)



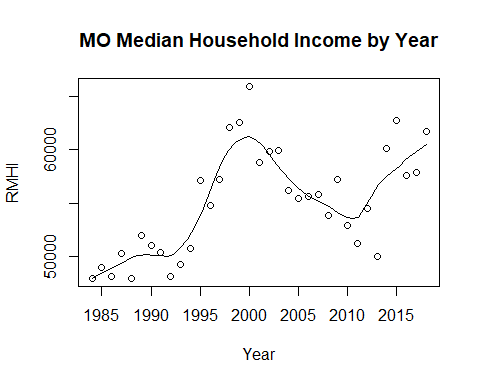
scatter.smooth(MO\_data$year, MO\_data$HPI, main="MO HPI by Year", xlab = "Year", ylab = "HPI", span = 2/8)



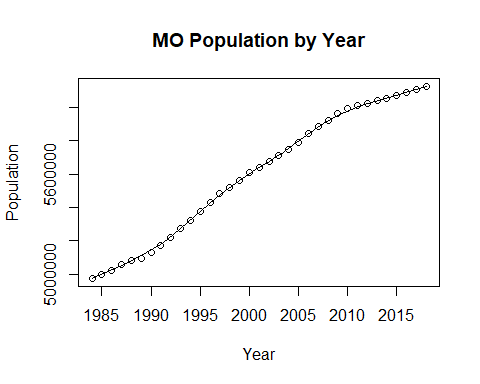
scatter.smooth(x=MO\_data$year, y=MO\_data$poverty, main="MO % of People Below Poverty Level", xlab = "Year", ylab = "Poverty", ylim = c(0,20), span = 2/8)



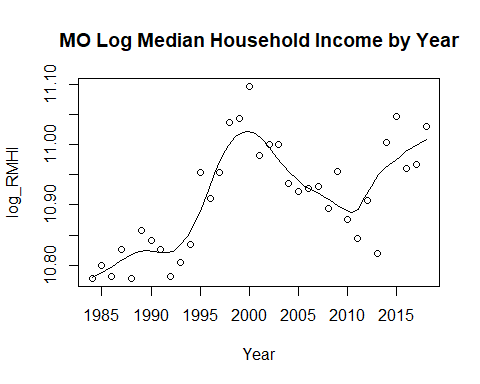
MO\_data$RMHI <- as.integer(MO\_data$RMHI)  
scatter.smooth(x=MO\_data$year, y=MO\_data$RMHI, main="MO Median Household Income by Year", xlab = "Year", ylab = "RMHI", span = 2/8 )



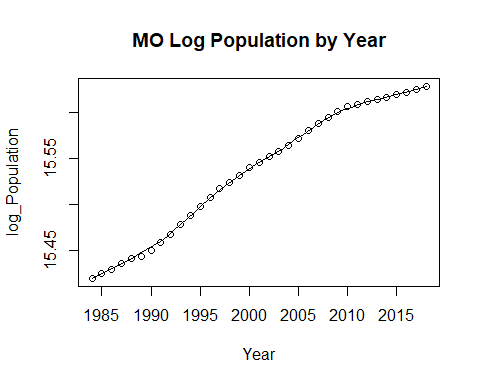
MO\_data$population <- as.integer(MO\_data$population)  
scatter.smooth(x=MO\_data$year, y=MO\_data$population, main="MO Population by Year", xlab = "Year", ylab = "Population", span = 2/8 )



#MO\_data$log\_RMHI <- as.integer(MO\_data$log\_RMHI)  
scatter.smooth(x=MO\_data$year, y=MO\_data$log\_RMHI, main="MO Log Median Household Income by Year", xlab = "Year", ylab = "log\_RMHI", span = 2/8 )



#MO\_data$log\_pop <- as.integer(MO\_data$log\_pop)  
scatter.smooth(x=MO\_data$year, y=MO\_data$log\_pop, main="MO Log Population by Year", xlab = "Year", ylab = "log\_Population", span = 2/8)



cor(MO\_data$RMHI, MO\_data$value)

## [1] -0.4832073

cor(MO\_data$HPI, MO\_data$value)

## [1] 0.05064891

cor(MO\_data$poverty, MO\_data$value)

## [1] 0.603927

cor(MO\_data$population, MO\_data$value)

## [1] 0.08687034

cor(MO\_data$sp500, MO\_data$value)

## [1] -0.09530716

cor(MO\_data$log\_pop, MO\_data$value)

## [1] 0.08057646

cor(MO\_data$log\_RMHI, MO\_data$value)

## [1] -0.4725887

#run base correlations between categories and dependent variable

THe correlation between the median household income and unemployment in Missouri is -0.4832073. The correlation between the median household income and unemployment is that of a medium strength negative relationship. A negative relationship between the two variables suggests that the two variables move inversely of one another with a medium stength relationship.

The correlation between the housing price index and unemployment in Missouri is 0.05064891. The correlation between the housing price index and unemployment is that of a small positive relationship. The relationship between the two variables is quite small with it being so close to 0, but does suggest that the two variables do move in the same direction of one another.

The correlation between poverty, the proportion of the population below the poverty line, and unemployment in Missouri is 0.603927. The correlation between poverty and unemployment is that of a large positive correlation. A large positive correlation suggests that the two variables tend to mimic each other. As poverty increases so does the unemployment rate.

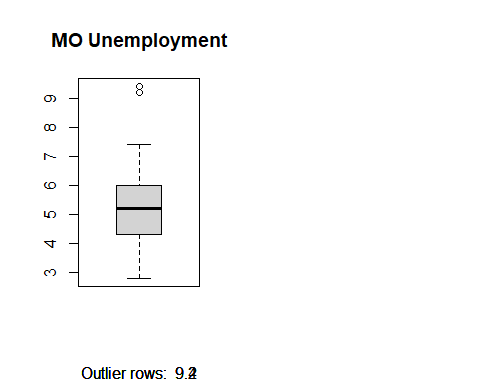
The correlation between the population and unemployment in Missouri is 0.08687034. The correlation between the population and unemployment is a small positive correlation. The small positive correlation suggests a weak relationship between the two variables, but does suggest that they move in the same direction of one another.

The correlation between the S&P 500 and unemployment in Missouri is -0.09530716. The correlation between the S&P 500 and unemployment is that of a small negative correlation. THe relationship between the two variables is a weak relationship, but does suggest that they move inversely of one another.

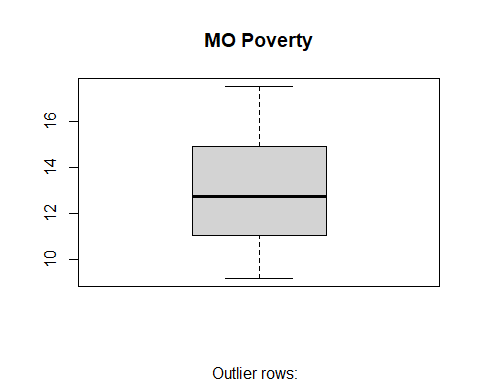
The correlation of the log of the population and unemployment in Missouri is 0.08057646. The correlation between the log of the population and unemployment is that of a small positive relationship. The stregth of the correlation is quite weak, but does suggest that the two variables move in the same direction of one another.

The correlation of the log of the median household income and unemployment in Missouri is -0.4725887. The correlation between the log of the median household income and unemployment suggest a medium positive correlation. The relationship between the two variables is a medium positive relationship suggesting the two variables tend to move in the same direction of one another.

par(mfrow=c(1, 2)) # divide graph area in 2 columns  
boxplot(MO\_data$value, main="MO Unemployment", sub=paste("Outlier rows: ", boxplot.stats(MO\_data$value)$out)) # box plot for 'Unemployment'



boxplot(MO\_data$poverty, main="MO Poverty", sub=paste("Outlier rows: ", boxplot.stats(MO\_data$poverty)$out)) # box plot for 'Poverty'



#create box plot for poverty level and unemployment level

#run multiple linear model for data  
MO\_reg1 <- lm(value ~ poverty + RMHI + HPI + population + sp500, data = MO\_data)  
summary(MO\_reg1)

##   
## Call:  
## lm(formula = value ~ poverty + RMHI + HPI + population + sp500,   
## data = MO\_data)  
##   
## Residuals:  
## Min 1Q Median 3Q Max   
## -1.9589 -0.5068 -0.0289 0.6061 3.5225   
##   
## Coefficients:  
## Estimate Std. Error t value Pr(>|t|)   
## (Intercept) -4.080e+00 1.058e+01 -0.386 0.7025   
## poverty 1.804e-01 1.800e-01 1.002 0.3244   
## RMHI -1.671e-04 9.667e-05 -1.729 0.0941 .  
## HPI -1.114e-02 1.502e-02 -0.742 0.4640   
## population 3.384e-06 2.692e-06 1.257 0.2185   
## sp500 -1.413e-02 1.449e-02 -0.975 0.3372   
## ---  
## Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1  
##   
## Residual standard error: 1.295 on 30 degrees of freedom  
## Multiple R-squared: 0.4573, Adjusted R-squared: 0.3669   
## F-statistic: 5.056 on 5 and 30 DF, p-value: 0.001772

anova(MO\_reg1)

## Analysis of Variance Table  
##   
## Response: value  
## Df Sum Sq Mean Sq F value Pr(>F)   
## poverty 1 33.821 33.821 20.1620 9.784e-05 \*\*\*  
## RMHI 1 0.251 0.251 0.1496 0.7017   
## HPI 1 4.813 4.813 2.8692 0.1006   
## population 1 1.925 1.925 1.1477 0.2926   
## sp500 1 1.596 1.596 0.9512 0.3372   
## Residuals 30 50.324 1.677   
## ---  
## Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1

MO\_reg2 <- lm(value ~ HPI + poverty + sp500 + log\_pop + log\_RMHI, data = MO\_data)  
summary(MO\_reg2)

##   
## Call:  
## lm(formula = value ~ HPI + poverty + sp500 + log\_pop + log\_RMHI,   
## data = MO\_data)  
##   
## Residuals:  
## Min 1Q Median 3Q Max   
## -1.9798 -0.5076 -0.0322 0.6245 3.5256   
##   
## Coefficients:  
## Estimate Std. Error t value Pr(>|t|)  
## (Intercept) -188.08253 208.41595 -0.902 0.374  
## HPI -0.01103 0.01480 -0.746 0.462  
## poverty 0.18321 0.18365 0.998 0.326  
## sp500 -0.01366 0.01451 -0.941 0.354  
## log\_pop 19.08100 14.90902 1.280 0.210  
## log\_RMHI -9.41671 5.61354 -1.677 0.104  
##   
## Residual standard error: 1.298 on 30 degrees of freedom  
## Multiple R-squared: 0.4547, Adjusted R-squared: 0.3638   
## F-statistic: 5.003 on 5 and 30 DF, p-value: 0.001891

anova(MO\_reg2)

## Analysis of Variance Table  
##   
## Response: value  
## Df Sum Sq Mean Sq F value Pr(>F)   
## HPI 1 0.238 0.238 0.1411 0.7098   
## poverty 1 35.349 35.349 20.9714 7.618e-05 \*\*\*  
## sp500 1 1.210 1.210 0.7176 0.4036   
## log\_pop 1 0.623 0.623 0.3696 0.5478   
## log\_RMHI 1 4.743 4.743 2.8140 0.1038   
## Residuals 30 50.567 1.686   
## ---  
## Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1

MO\_linearModelSignificant <- lm(value ~ poverty + log\_pop, data = MO\_data)  
summary(MO\_linearModelSignificant)

##   
## Call:  
## lm(formula = value ~ poverty + log\_pop, data = MO\_data)  
##   
## Residuals:  
## Min 1Q Median 3Q Max   
## -2.53194 -0.61245 0.04034 0.62686 2.58796   
##   
## Coefficients:  
## Estimate Std. Error t value Pr(>|t|)   
## (Intercept) -53.81840 49.87591 -1.079 0.288   
## poverty 0.44395 0.09824 4.519 7.55e-05 \*\*\*  
## log\_pop 3.44541 3.20012 1.077 0.289   
## ---  
## Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1  
##   
## Residual standard error: 1.313 on 33 degrees of freedom  
## Multiple R-squared: 0.3863, Adjusted R-squared: 0.3491   
## F-statistic: 10.39 on 2 and 33 DF, p-value: 0.0003173

WA\_data <- merged\_final[ merged\_final$state == "WA", ]

#Descriptive statistics  
stat.desc(WA\_data)

## year period periodName value state quarter annual  
## nbr.val 3.600000e+01 NA NA 36.0000000 NA 36 36  
## nbr.null 0.000000e+00 NA NA 0.0000000 NA 0 0  
## nbr.na 0.000000e+00 NA NA 0.0000000 NA 0 0  
## min 1.984000e+03 NA NA 4.5000000 NA 4 1  
## max 2.018000e+03 NA NA 10.5000000 NA 4 1  
## range 3.400000e+01 NA NA 6.0000000 NA 0 0  
## sum 7.204500e+04 NA NA 235.6000000 NA 144 36  
## median 2.001500e+03 NA NA 6.2500000 NA 4 1  
## mean 2.001250e+03 NA NA 6.5444444 NA 4 1  
## SE.mean 1.701715e+00 NA NA 0.2642883 NA 0 0  
## CI.mean.0.95 3.454665e+00 NA NA 0.5365338 NA 0 0  
## var 1.042500e+02 NA NA 2.5145397 NA 0 0  
## std.dev 1.021029e+01 NA NA 1.5857300 NA 0 0  
## coef.var 5.101956e-03 NA NA 0.2423017 NA 0 0  
## HPI MHI RMHI poverty population  
## nbr.val 3.600000e+01 3.600000e+01 3.600000e+01 36.0000000 3.600000e+01  
## nbr.null 0.000000e+00 0.000000e+00 0.000000e+00 0.0000000 0.000000e+00  
## nbr.na 0.000000e+00 0.000000e+00 0.000000e+00 0.0000000 0.000000e+00  
## min 1.101900e+02 2.400000e+04 5.356200e+04 8.0000000 4.343656e+06  
## max 6.329200e+02 7.972600e+04 7.972600e+04 12.9000000 7.535591e+06  
## range 5.227300e+02 5.572600e+04 2.616400e+04 4.9000000 3.191935e+06  
## sum 1.143666e+04 1.697647e+06 2.333727e+06 390.4000000 2.143828e+08  
## median 3.012150e+02 4.644700e+04 6.482200e+04 11.1000000 6.019036e+06  
## mean 3.176850e+02 4.715686e+04 6.482575e+04 10.8444444 5.955079e+06  
## SE.mean 2.419525e+01 2.429641e+03 9.611473e+02 0.2160778 1.565310e+05  
## CI.mean.0.95 4.911897e+01 4.932434e+03 1.951233e+03 0.4386612 3.177747e+05  
## var 2.107476e+04 2.125137e+08 3.325695e+07 1.6808254 8.820699e+11  
## std.dev 1.451715e+02 1.457785e+04 5.766884e+03 1.2964665 9.391857e+05  
## coef.var 4.569668e-01 3.091353e-01 8.895977e-02 0.1195512 1.577117e-01  
## sp500 log\_pop log\_RMHI  
## nbr.val 36.000000 36.00000000 3.600000e+01  
## nbr.null 1.000000 0.00000000 0.000000e+00  
## nbr.na 0.000000 0.00000000 0.000000e+00  
## min -38.490000 15.28422695 1.088860e+01  
## max 34.110000 15.83514782 1.128635e+01  
## range 72.600000 0.55092087 3.977559e-01  
## sum 341.220000 561.13786027 3.987234e+02  
## median 11.895000 15.61042227 1.107940e+01  
## mean 9.478333 15.58716279 1.107565e+01  
## SE.mean 2.652485 0.02712181 1.472286e-02  
## CI.mean.0.95 5.384831 0.05506021 2.988900e-02  
## var 253.284340 0.02648134 7.803454e-03  
## std.dev 15.914909 0.16273089 8.833716e-02  
## coef.var 1.679083 0.01044006 7.975799e-03

The dependent variable, unemployment, has a mean of 6.54 and a standard deviation of 1.59. The lowest unemployment rate was in 1997 at 4.5, while the highest was in 2009 at 10.5. The mean HPI was 317.69, with a standard deviation of 145.17. The lowest HPI was in 1984 at 110.19, while the highest was in 2018 at 632.92. Poverty rate had a mean of 10.84 and a standard deviation of 1.30. In 2006, lowest rate was seen of 8.0, and the highest was 12.9 in 1986. The sp500 index mean was 9.48, and the standard deviation was 15.70. In 2008, the lowest index was seen at -38.4, while the highest index was seen in 1995 at 34.11. The logged population mean was 15.59, and the standard deviation was 0.16. The lowest logged population at 15.28 in 1984, and the highest in 2018 at 15.84. Logged real median household income had a mean of 11.08 and a standard deviation of 0.088. In 1985, Washington had the lowest log\_RMHI at 10.89, while had the highest in 2018 at 11.29.

#pairwise correlations  
sapply(WA\_data, class)

## year period periodName value state quarter   
## "numeric" "character" "character" "numeric" "character" "numeric"   
## annual HPI MHI RMHI poverty population   
## "numeric" "numeric" "numeric" "numeric" "numeric" "numeric"   
## sp500 log\_pop log\_RMHI   
## "numeric" "numeric" "numeric"

sapply(WA\_data, is.factor)

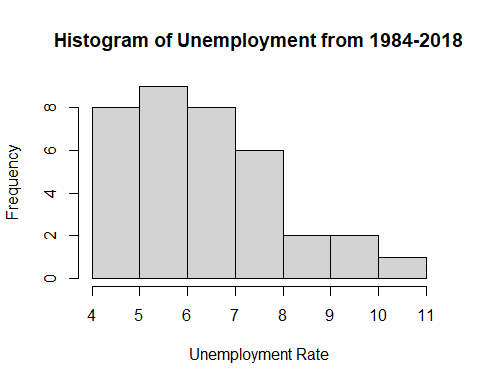
## year period periodName value state quarter annual   
## FALSE FALSE FALSE FALSE FALSE FALSE FALSE   
## HPI MHI RMHI poverty population sp500 log\_pop   
## FALSE FALSE FALSE FALSE FALSE FALSE FALSE   
## log\_RMHI   
## FALSE

cor(WA\_data[sapply(WA\_data, function(x) !is.character(x))])

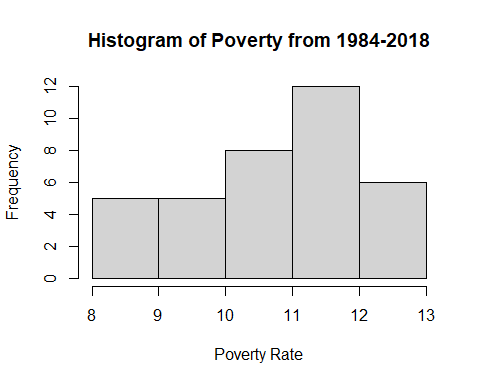
## Warning in cor(WA\_data[sapply(WA\_data, function(x) !is.character(x))]): the  
## standard deviation is zero

## year value quarter annual HPI MHI  
## year 1.00000000 -0.09846885 NA NA 0.95497907 0.97834561  
## value -0.09846885 1.00000000 NA NA -0.19784393 -0.16995997  
## quarter NA NA 1 NA NA NA  
## annual NA NA NA 1 NA NA  
## HPI 0.95497907 -0.19784393 NA NA 1.00000000 0.96373904  
## MHI 0.97834561 -0.16995997 NA NA 0.96373904 1.00000000  
## RMHI 0.73244639 -0.38197496 NA NA 0.76924404 0.84445113  
## poverty 0.09475391 0.55199926 NA NA -0.05179519 -0.02984112  
## population 0.99760675 -0.11590724 NA NA 0.95242097 0.97597644  
## sp500 -0.13852928 -0.02637235 NA NA -0.20212589 -0.10232708  
## log\_pop 0.99167165 -0.12589985 NA NA 0.94330829 0.96393083  
## log\_RMHI 0.73798223 -0.37516803 NA NA 0.76933961 0.84503049  
## RMHI poverty population sp500 log\_pop  
## year 0.73244639 0.09475391 0.9976068 -0.13852928 0.9916716  
## value -0.38197496 0.55199926 -0.1159072 -0.02637235 -0.1258998  
## quarter NA NA NA NA NA  
## annual NA NA NA NA NA  
## HPI 0.76924404 -0.05179519 0.9524210 -0.20212589 0.9433083  
## MHI 0.84445113 -0.02984112 0.9759764 -0.10232708 0.9639308  
## RMHI 1.00000000 -0.38393686 0.7366598 0.04794022 0.7271869  
## poverty -0.38393686 1.00000000 0.0882941 0.10971620 0.0842142  
## population 0.73665979 0.08829410 1.0000000 -0.13576575 0.9971539  
## sp500 0.04794022 0.10971620 -0.1357657 1.00000000 -0.1370012  
## log\_pop 0.72718694 0.08421420 0.9971539 -0.13700122 1.0000000  
## log\_RMHI 0.99852049 -0.38047987 0.7416809 0.04551562 0.7343848  
## log\_RMHI  
## year 0.73798223  
## value -0.37516803  
## quarter NA  
## annual NA  
## HPI 0.76933961  
## MHI 0.84503049  
## RMHI 0.99852049  
## poverty -0.38047987  
## population 0.74168090  
## sp500 0.04551562  
## log\_pop 0.73438477  
## log\_RMHI 1.00000000

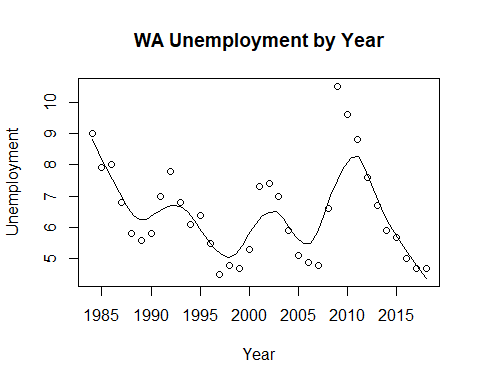
hist(WA\_data$value,  
 main="Histogram of Unemployment from 1984-2018",  
 xlab="Unemployment Rate",  
 breaks = 5)

 In Washington, from 1984-2018,the annual unemployment rates ranged from 4.0-10.9%. The histogram displays a positively skewed distribution meaning relatively low poverty rates occurred more than higher rates. There is a small presence of higher poverty rates and only 1 year that experienced a poverty rate above 10%.

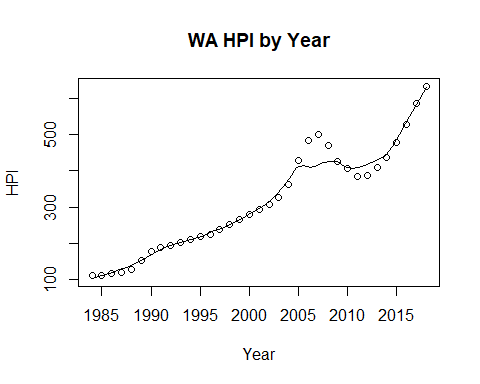
hist(WA\_data$poverty,  
 main="Histogram of Poverty from 1984-2018",  
 xlab="Poverty Rate",  
 breaks = 5)

 In Washington, from 1984 to 2018, The annual poverty rates ranged from 8.0-12.9%. The histogram appears to be slightly negatively skewed. 18 of the 36 years were characterized by poverty rates the occurred in the two highest intervals, 11.0-11.9% and 12.0-12.9%.

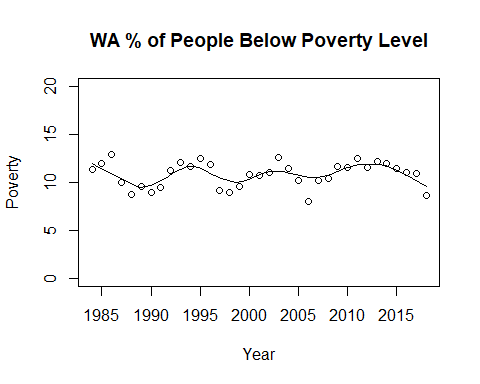
scatter.smooth(WA\_data$year, WA\_data$value, main="WA Unemployment by Year", xlab = "Year", ylab = "Unemployment", span = 2/8)



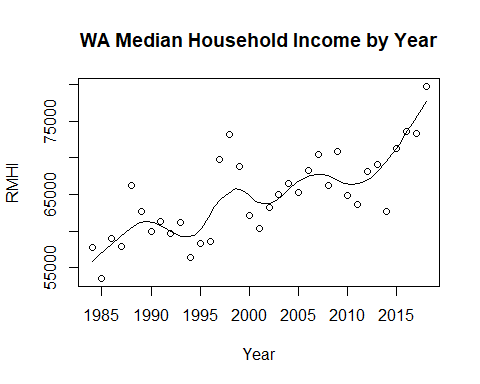
scatter.smooth(WA\_data$year, WA\_data$HPI, main="WA HPI by Year", xlab = "Year", ylab = "HPI", span = 2/8)



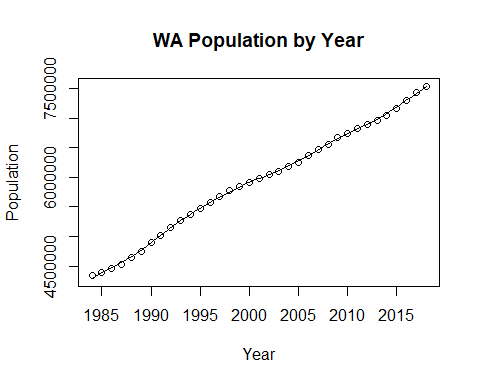
scatter.smooth(x=WA\_data$year, y=WA\_data$poverty, main="WA % of People Below Poverty Level", xlab = "Year", ylab = "Poverty", ylim = c(0,20), span = 2/8)



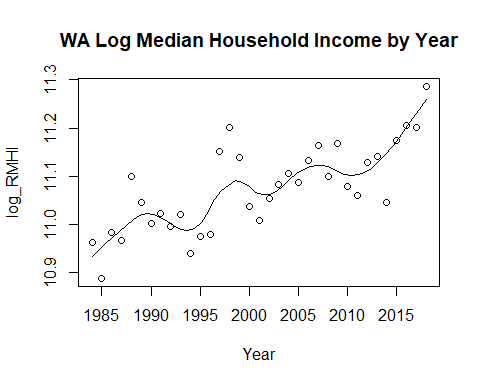
WA\_data$RMHI <- as.integer(WA\_data$RMHI)  
scatter.smooth(x=WA\_data$year, y=WA\_data$RMHI, main="WA Median Household Income by Year", xlab = "Year", ylab = "RMHI", span = 2/8 )



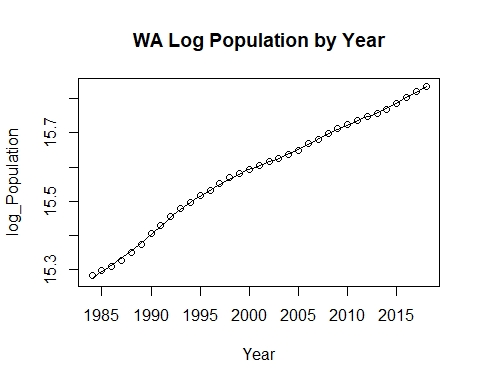
WA\_data$population <- as.integer(WA\_data$population)  
scatter.smooth(x=WA\_data$year, y=WA\_data$population, main="WA Population by Year", xlab = "Year", ylab = "Population", span = 2/8 )



#WA\_data$log\_RMHI <- as.integer(WA\_data$log\_RMHI)  
scatter.smooth(x=WA\_data$year, y=WA\_data$log\_RMHI, main="WA Log Median Household Income by Year", xlab = "Year", ylab = "log\_RMHI", span = 2/8 )



#WA\_data$log\_pop <- as.integer(WA\_data$log\_pop)  
scatter.smooth(x=WA\_data$year, y=WA\_data$log\_pop, main="WA Log Population by Year", xlab = "Year", ylab = "log\_Population", span = 2/8)



cor(WA\_data$RMHI, WA\_data$value)

## [1] -0.381975

cor(WA\_data$HPI, WA\_data$value)

## [1] -0.1978439

cor(WA\_data$poverty, WA\_data$value)

## [1] 0.5519993

cor(WA\_data$population, WA\_data$value)

## [1] -0.1159072

cor(WA\_data$sp500, WA\_data$value)

## [1] -0.02637235

cor(WA\_data$log\_pop, WA\_data$value)

## [1] -0.1258998

cor(WA\_data$log\_RMHI, WA\_data$value)

## [1] -0.375168

#run base correlations between categories and dependent variable

The correlation between the median household income and unemployment in Washington is -0.381975. The correlation between the median household income and unemployment is that of a medium negative correlation. The relationship between the two variables has medium strength and suggests that they move inversely of one another.

The correlation between the housing price index and unemployment in Washington is -0.1978439. The correlation between the housing price index and unemployment is that of a small negative correlation. A small negative correlation between the two variables suggests a weak inverse relationship between the housing price index and unemployment.

The correlation between poverty, the proportion of the population below the poverty line, and unemployment in Washington is 0.5519993. The correlation between poverty and unemployment is that of a large positive correlation. A large positive correlation suggests a strong positive relationship between the two variables that tend to move in the direction of one another.

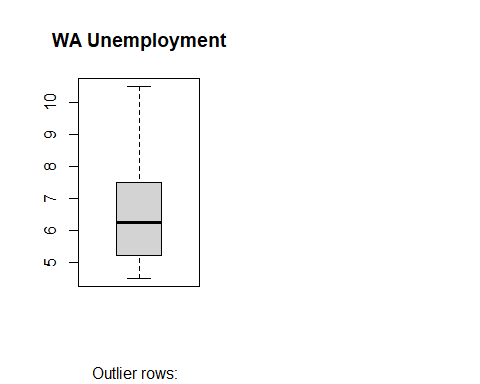
The correlation between the population and unemployment in Washington is -0.1159072. The correlation between the population and unemployment is that of a small negative correlation. A small negative correlation suggests a weak inverse relationship between the two variables.

The correlation between the S&P 500 and unemployment in Washington is -0.02637235. The correlation between the S&P 500 and unemployment is that of a small negative correlation. A small negative correlation suggests a weak inverse relationship between the two variables.

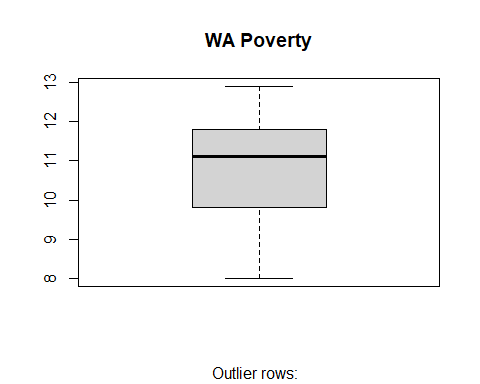
The correlation between the log of the population and unemployment in Washington is -0.1258998. The correlation between the log of the population and unemployment is that of a small negative correlation. The small negative correlation suggests a weak inverse relationship between the two variables.

The correlation between the log of the median household income and unemployment in Washington is -0.375168. THe correlation between the log of the median household income and unemployment is that of a medium negative Correlation. A medium negative correlation suggests a somewhat strong inverse relationship between the two variables.

par(mfrow=c(1, 2)) # divide graph area in 2 columns  
boxplot(WA\_data$value, main="WA Unemployment", sub=paste("Outlier rows: ", boxplot.stats(WA\_data$value)$out)) # box plot for 'Unemployment'



boxplot(WA\_data$poverty, main="WA Poverty", sub=paste("Outlier rows: ", boxplot.stats(WA\_data$poverty)$out)) # box plot for 'Poverty'



#create box plot for poverty level and unemployment level

#run multiple linear model for data  
WA\_reg1 <- lm(value ~ poverty + RMHI + HPI + population + sp500, data = WA\_data)  
summary(WA\_reg1)

##   
## Call:  
## lm(formula = value ~ poverty + RMHI + HPI + population + sp500,   
## data = WA\_data)  
##   
## Residuals:  
## Min 1Q Median 3Q Max   
## -1.8810 -0.9367 -0.1978 0.4889 3.8449   
##   
## Coefficients:  
## Estimate Std. Error t value Pr(>|t|)   
## (Intercept) 9.431e-01 6.443e+00 0.146 0.8846   
## poverty 6.342e-01 2.683e-01 2.364 0.0248 \*  
## RMHI -1.937e-05 8.897e-05 -0.218 0.8292   
## HPI -2.405e-03 6.100e-03 -0.394 0.6961   
## population 1.428e-07 9.769e-07 0.146 0.8847   
## sp500 -1.125e-02 1.689e-02 -0.666 0.5105   
## ---  
## Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1  
##   
## Residual standard error: 1.381 on 30 degrees of freedom  
## Multiple R-squared: 0.3498, Adjusted R-squared: 0.2415   
## F-statistic: 3.228 on 5 and 30 DF, p-value: 0.01891

anova(WA\_reg1)

## Analysis of Variance Table  
##   
## Response: value  
## Df Sum Sq Mean Sq F value Pr(>F)   
## poverty 1 26.817 26.8166 14.0597 0.0007561 \*\*\*  
## RMHI 1 2.985 2.9847 1.5648 0.2206262   
## HPI 1 0.102 0.1019 0.0534 0.8188002   
## population 1 0.039 0.0395 0.0207 0.8865625   
## sp500 1 0.846 0.8461 0.4436 0.5104803   
## Residuals 30 57.220 1.9073   
## ---  
## Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1

WA\_reg2 <- lm(value ~ HPI + poverty + sp500 + log\_pop + log\_RMHI, data = WA\_data)  
summary(WA\_reg2)

##   
## Call:  
## lm(formula = value ~ HPI + poverty + sp500 + log\_pop + log\_RMHI,   
## data = WA\_data)  
##   
## Residuals:  
## Min 1Q Median 3Q Max   
## -1.8267 -0.9406 -0.2529 0.5597 3.7919   
##   
## Coefficients:  
## Estimate Std. Error t value Pr(>|t|)   
## (Intercept) 13.451038 82.870199 0.162 0.8721   
## HPI -0.001297 0.005503 -0.236 0.8153   
## poverty 0.678607 0.258839 2.622 0.0136 \*  
## sp500 -0.011788 0.016833 -0.700 0.4892   
## log\_pop -0.579274 5.000767 -0.116 0.9086   
## log\_RMHI -0.425511 5.739510 -0.074 0.9414   
## ---  
## Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1  
##   
## Residual standard error: 1.382 on 30 degrees of freedom  
## Multiple R-squared: 0.3492, Adjusted R-squared: 0.2408   
## F-statistic: 3.22 on 5 and 30 DF, p-value: 0.01913

anova(WA\_reg2)

## Analysis of Variance Table  
##   
## Response: value  
## Df Sum Sq Mean Sq F value Pr(>F)   
## HPI 1 3.445 3.4449 1.8045 0.1892485   
## poverty 1 25.900 25.8997 13.5665 0.0009045 \*\*\*  
## sp500 1 1.341 1.3412 0.7025 0.4085697   
## log\_pop 1 0.040 0.0402 0.0211 0.8855836   
## log\_RMHI 1 0.010 0.0105 0.0055 0.9413934   
## Residuals 30 57.273 1.9091   
## ---  
## Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1

WA\_linearModelSignificant <- lm(value ~ poverty + log\_pop, data = WA\_data)  
summary(WA\_linearModelSignificant)

##   
## Call:  
## lm(formula = value ~ poverty + log\_pop, data = WA\_data)  
##   
## Residuals:  
## Min 1Q Median 3Q Max   
## -1.8679 -0.9720 -0.4343 0.7903 3.5751   
##   
## Coefficients:  
## Estimate Std. Error t value Pr(>|t|)   
## (Intercept) 25.3993 21.5688 1.178 0.24738   
## poverty 0.6930 0.1743 3.976 0.00036 \*\*\*  
## log\_pop -1.6918 1.3886 -1.218 0.23173   
## ---  
## Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1  
##   
## Residual standard error: 1.332 on 33 degrees of freedom  
## Multiple R-squared: 0.3346, Adjusted R-squared: 0.2943   
## F-statistic: 8.298 on 2 and 33 DF, p-value: 0.001204

OR\_data <- merged\_final[ merged\_final$state == "OR", ]

#Descriptive statistics  
stat.desc(OR\_data)

## year period periodName value state quarter annual  
## nbr.val 3.600000e+01 NA NA 36.0000000 NA 36 36  
## nbr.null 0.000000e+00 NA NA 0.0000000 NA 0 0  
## nbr.na 0.000000e+00 NA NA 0.0000000 NA 0 0  
## min 1.984000e+03 NA NA 3.9000000 NA 4 1  
## max 2.018000e+03 NA NA 10.8000000 NA 4 1  
## range 3.400000e+01 NA NA 6.9000000 NA 0 0  
## sum 7.204500e+04 NA NA 232.4000000 NA 144 36  
## median 2.001500e+03 NA NA 5.7500000 NA 4 1  
## mean 2.001250e+03 NA NA 6.4555556 NA 4 1  
## SE.mean 1.701715e+00 NA NA 0.3159879 NA 0 0  
## CI.mean.0.95 3.454665e+00 NA NA 0.6414894 NA 0 0  
## var 1.042500e+02 NA NA 3.5945397 NA 0 0  
## std.dev 1.021029e+01 NA NA 1.8959271 NA 0 0  
## coef.var 5.101956e-03 NA NA 0.2936892 NA 0 0  
## HPI MHI RMHI poverty population  
## nbr.val 3.600000e+01 3.600000e+01 3.600000e+01 36.0000000 3.600000e+01  
## nbr.null 0.000000e+00 0.000000e+00 0.000000e+00 0.0000000 0.000000e+00  
## nbr.na 0.000000e+00 0.000000e+00 0.000000e+00 0.0000000 0.000000e+00  
## min 9.718000e+01 2.139900e+04 4.886200e+04 9.2000000 2.666588e+06  
## max 5.666300e+02 6.916500e+04 6.916500e+04 15.0000000 4.190713e+06  
## range 4.694500e+02 4.776600e+04 2.030300e+04 5.8000000 1.524125e+06  
## sum 1.036012e+04 1.508700e+06 2.081811e+06 441.0000000 1.235918e+08  
## median 2.746050e+02 4.145550e+04 5.762800e+04 11.8500000 3.490681e+06  
## mean 2.877811e+02 4.190833e+04 5.782808e+04 12.2500000 3.433106e+06  
## SE.mean 2.291038e+01 2.034774e+03 6.838691e+02 0.2332483 7.904400e+04  
## CI.mean.0.95 4.651054e+01 4.130811e+03 1.388328e+03 0.4735192 1.604679e+05  
## var 1.889588e+04 1.490510e+08 1.683637e+07 1.9585714 2.249264e+11  
## std.dev 1.374623e+02 1.220864e+04 4.103215e+03 1.3994897 4.742640e+05  
## coef.var 4.776626e-01 2.913178e-01 7.095540e-02 0.1142441 1.381443e-01  
## sp500 log\_pop log\_RMHI  
## nbr.val 36.000000 3.600000e+01 3.600000e+01  
## nbr.null 1.000000 0.000000e+00 0.000000e+00  
## nbr.na 0.000000 0.000000e+00 0.000000e+00  
## min -38.490000 1.479631e+01 1.079676e+01  
## max 34.110000 1.524838e+01 1.114425e+01  
## range 72.600000 4.520711e-01 3.474950e-01  
## sum 341.220000 5.414165e+02 3.946603e+02  
## median 11.895000 1.506559e+01 1.096176e+01  
## mean 9.478333 1.503935e+01 1.096279e+01  
## SE.mean 2.652485 2.368634e-02 1.182095e-02  
## CI.mean.0.95 5.384831 4.808582e-02 2.399781e-02  
## var 253.284340 2.019753e-02 5.030459e-03  
## std.dev 15.914909 1.421180e-01 7.092573e-02  
## coef.var 1.679083 9.449747e-03 6.469681e-03

The dependent variable, unemployment, has a mean of 6.46 and a standard deviation of 1.90. The lowest unemployment rate was in 2017 at 3.9, while the highest was in 2009 at 10.8. The mean HPI was 287.78, with a standard deviation of 137.46. The lowest HPI was in 1985 at 97.18, while the highest was in 2018 at 566.63. Poverty rate had a mean of 12.25 and a standard deviation of 1.40. In 1990, lowest rate was seen of 9.2, and the highest was 15.0 in 1998. The sp500 index mean was 9.48, and the standard deviation was 15.70. In 2008, the lowest index was seen at -38.4, while the highest index was seen in 1995 at 34.11. The logged population mean was 15.04, and the standard deviation was 0.14. The lowest logged population at 14.796 in 1984, and the highest in 2018 at 15.25. Logged real median household income had a mean of 10.96 and a standard deviation of 0.071. In 1985, Oregon had the lowest log\_RMHI at 10.80, while had the highest in 2018 at 11.14.

#pairwise correlations  
sapply(OR\_data, class)

## year period periodName value state quarter   
## "numeric" "character" "character" "numeric" "character" "numeric"   
## annual HPI MHI RMHI poverty population   
## "numeric" "numeric" "numeric" "numeric" "numeric" "numeric"   
## sp500 log\_pop log\_RMHI   
## "numeric" "numeric" "numeric"

sapply(OR\_data, is.factor)

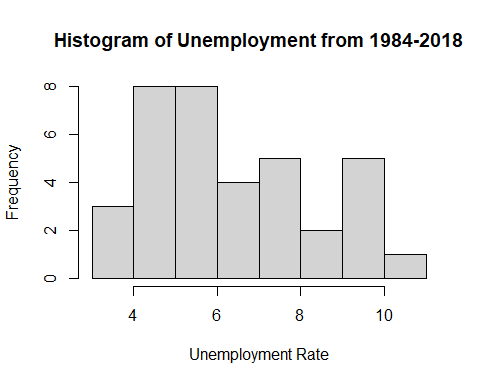
## year period periodName value state quarter annual   
## FALSE FALSE FALSE FALSE FALSE FALSE FALSE   
## HPI MHI RMHI poverty population sp500 log\_pop   
## FALSE FALSE FALSE FALSE FALSE FALSE FALSE   
## log\_RMHI   
## FALSE

cor(OR\_data[sapply(OR\_data, function(x) !is.character(x))])

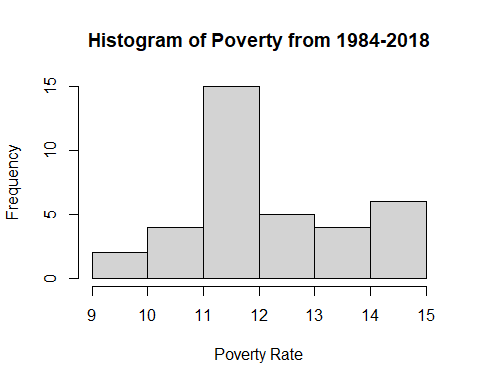
## Warning in cor(OR\_data[sapply(OR\_data, function(x) !is.character(x))]): the  
## standard deviation is zero

## year value quarter annual HPI MHI  
## year 1.00000000 0.02612435 NA NA 0.95912382 0.98127265  
## value 0.02612435 1.00000000 NA NA -0.07066041 -0.06421313  
## quarter NA NA 1 NA NA NA  
## annual NA NA NA 1 NA NA  
## HPI 0.95912382 -0.07066041 NA NA 1.00000000 0.96237600  
## MHI 0.98127265 -0.06421313 NA NA 0.96237600 1.00000000  
## RMHI 0.68937724 -0.39509776 NA NA 0.71729570 0.80057618  
## poverty 0.18685448 0.44773917 NA NA 0.06569451 0.09803790  
## population 0.99587797 0.03283668 NA NA 0.96049606 0.97695728  
## sp500 -0.13852928 -0.09391935 NA NA -0.18376601 -0.18174097  
## log\_pop 0.99016952 0.03392033 NA NA 0.95319408 0.96821393  
## log\_RMHI 0.68950977 -0.38874524 NA NA 0.71276190 0.79665604  
## RMHI poverty population sp500 log\_pop  
## year 0.6893772 0.18685448 0.99587797 -0.13852928 0.99016952  
## value -0.3950978 0.44773917 0.03283668 -0.09391935 0.03392033  
## quarter NA NA NA NA NA  
## annual NA NA NA NA NA  
## HPI 0.7172957 0.06569451 0.96049606 -0.18376601 0.95319408  
## MHI 0.8005762 0.09803790 0.97695728 -0.18174097 0.96821393  
## RMHI 1.0000000 -0.17905708 0.69884405 -0.20659566 0.69830244  
## poverty -0.1790571 1.00000000 0.18411082 0.36989977 0.18507510  
## population 0.6988441 0.18411082 1.00000000 -0.14258053 0.99815701  
## sp500 -0.2065957 0.36989977 -0.14258053 1.00000000 -0.14294569  
## log\_pop 0.6983024 0.18507510 0.99815701 -0.14294569 1.00000000  
## log\_RMHI 0.9984524 -0.16869294 0.69973663 -0.20234683 0.70140009  
## log\_RMHI  
## year 0.6895098  
## value -0.3887452  
## quarter NA  
## annual NA  
## HPI 0.7127619  
## MHI 0.7966560  
## RMHI 0.9984524  
## poverty -0.1686929  
## population 0.6997366  
## sp500 -0.2023468  
## log\_pop 0.7014001  
## log\_RMHI 1.0000000

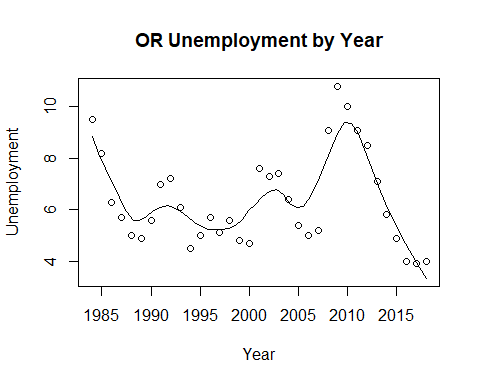
hist(OR\_data$value,  
 main="Histogram of Unemployment from 1984-2018",  
 xlab="Unemployment Rate",  
 breaks = 5)

 In Oregon, from 1984 to 2018, the annual unemployment rates ranged from 2.0%-11.9%.The histogram did not produce a distribution with a skew worth acknowledging, however, the two most frequent ranges were 4.0-4.9% & 5.0-5.9% consecutively. There was only one year with an unemployment rate of 10.0% or higher.

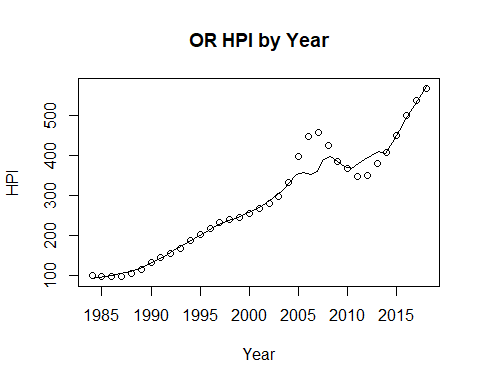
hist(OR\_data$poverty,  
 main="Histogram of Poverty from 1984-2018",  
 xlab="Poverty Rate",  
 breaks = 5)

 In Oregon, from 1984 to 2018, the annual unemployment rates ranged from 9.0%-14.9%. The most frequent range of unemployment rates was 11.0-11.9% which occurred 15 of the 36 years.

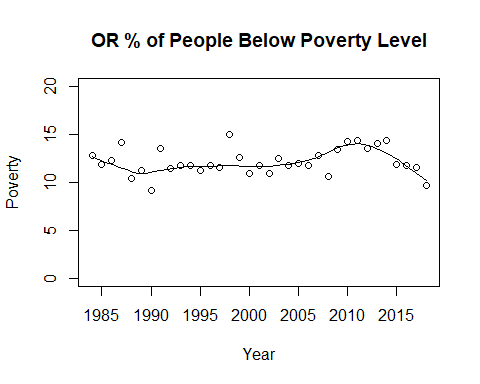
scatter.smooth(OR\_data$year, OR\_data$value, main="OR Unemployment by Year", xlab = "Year", ylab = "Unemployment", span = 2/8)



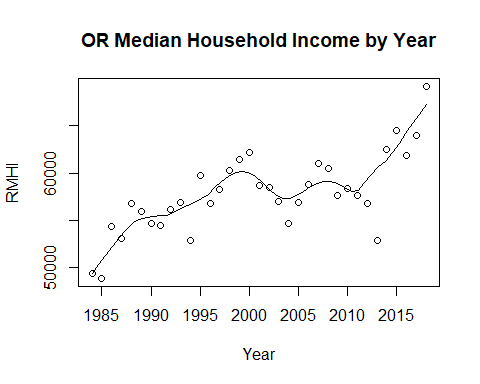
scatter.smooth(OR\_data$year, OR\_data$HPI, main="OR HPI by Year", xlab = "Year", ylab = "HPI", span = 2/8)



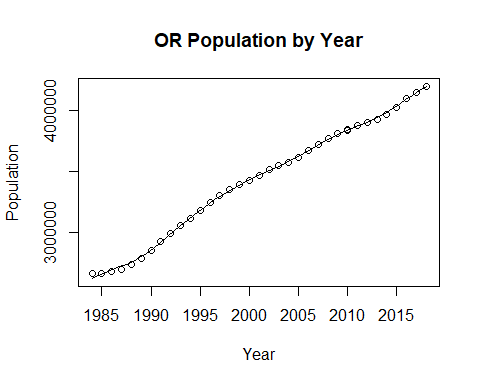
scatter.smooth(x=OR\_data$year, y=OR\_data$poverty, main="OR % of People Below Poverty Level", xlab = "Year", ylab = "Poverty", ylim = c(0,20), span = 2/8)



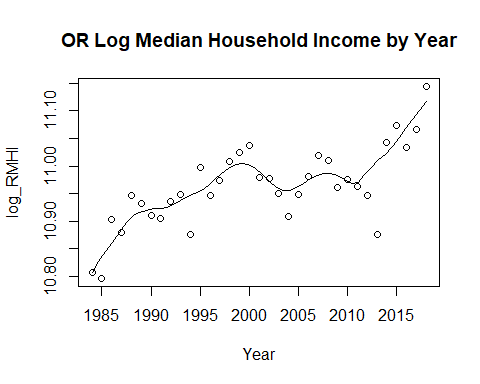
OR\_data$RMHI <- as.integer(OR\_data$RMHI)  
scatter.smooth(x=OR\_data$year, y=OR\_data$RMHI, main="OR Median Household Income by Year", xlab = "Year", ylab = "RMHI", span = 2/8 )



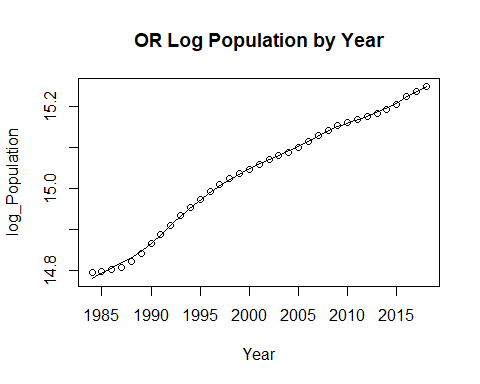
OR\_data$population <- as.integer(OR\_data$population)  
scatter.smooth(x=OR\_data$year, y=OR\_data$population, main="OR Population by Year", xlab = "Year", ylab = "Population", span = 2/8 )



#OR\_data$log\_RMHI <- as.integer(OR\_data$log\_RMHI)  
scatter.smooth(x=OR\_data$year, y=OR\_data$log\_RMHI, main="OR Log Median Household Income by Year", xlab = "Year", ylab = "log\_RMHI", span = 2/8 )



#OR\_data$log\_pop <- as.integer(OR\_data$log\_pop)  
scatter.smooth(x=OR\_data$year, y=OR\_data$log\_pop, main="OR Log Population by Year", xlab = "Year", ylab = "log\_Population", span = 2/8)



cor(OR\_data$RMHI, OR\_data$value)

## [1] -0.3950978

cor(OR\_data$HPI, OR\_data$value)

## [1] -0.07066041

cor(OR\_data$poverty, OR\_data$value)

## [1] 0.4477392

cor(OR\_data$population, OR\_data$value)

## [1] 0.03283668

cor(OR\_data$sp500, OR\_data$value)

## [1] -0.09391935

cor(OR\_data$log\_pop, OR\_data$value)

## [1] 0.03392033

cor(OR\_data$log\_RMHI, OR\_data$value)

## [1] -0.3887452

#run base correlations between categories and dependent variable

The correlation between the median household income and unemployment in Oregon is -0.3950978. The correlation between the median household income and unemployment is that of an inverse relationship. The inverse relationship is not the strongest, but does shows that in theory that as the median household income goes up, then the unemployment would go down. Of course, When it comes to scenarios such as a pandemic or depression, this could cause the relationship to become much more skewed, then it would be in most cases.

The correlation between the housing price index and unemployment in Oregon is -0.07066041. The correlation between the housing price index and unemployment is an extremely weak relationship. With the correlation being that close to 0, it shows that there is a possibility of their being no correlation between the two variables.

The correlation between poverty, the percentage of the population below the poverty line, and unemployment in Oregon is 0.4477392. The correlation between poverty and unemployment is a decently strong positive relationship. This correlation suggests that the two variables tend to head in the same direction. As one goes up the other tends to do the same.

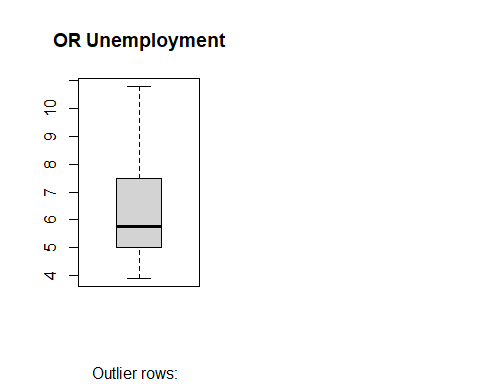
The correlation between the population and unemployment in Oregon is 0.03283668. The correlation between population and unemployment is a weak positive correlation. With the correlation being so close to 0, the correlation suggests that there is no relationship between the two variables.

The correlation between the S&P 500 and unemployment in Oregon is -0.09391935. The correlation between the S&P 500 and unemployment is a weak negative relationship. A negative correlation suggests an inverse relationship between the two variables. But, the inverse relationship is extremely weak due to the closeness of the correlation being to 0.

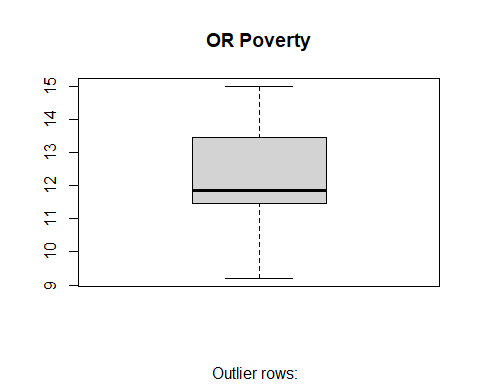
The correlation between the log of the population and unemployment in Oregon is 0.03392033. The correlation between the log of the population and unemployment is a weak positive correlation. With the correlation being so close to 0, it shows that relationship between the two variables is weak and does not exactly distinguish the direction that the two variables can go.

The correlation between the log of the median household income and unemployment in Oregon is -0.3887452. The correlation between the log of the median household income and unemployment is a seemingly significant negative relationship. The negative relationship of correlation suggests that the two variables tend to move in opposite directions of each other.

par(mfrow=c(1, 2)) # divide graph area in 2 columns  
boxplot(OR\_data$value, main="OR Unemployment", sub=paste("Outlier rows: ", boxplot.stats(OR\_data$value)$out)) # box plot for 'Unemployment'



boxplot(OR\_data$poverty, main="OR Poverty", sub=paste("Outlier rows: ", boxplot.stats(OR\_data$poverty)$out)) # box plot for 'Poverty'



#create box plot for poverty level and unemployment level

#run multiple linear model for data  
OR\_reg1 <- lm(value ~ poverty + RMHI + HPI + population + sp500, data = OR\_data)  
summary(OR\_reg1)

##   
## Call:  
## lm(formula = value ~ poverty + RMHI + HPI + population + sp500,   
## data = OR\_data)  
##   
## Residuals:  
## Min 1Q Median 3Q Max   
## -3.3111 -0.7164 -0.0853 0.7328 3.6921   
##   
## Coefficients:  
## Estimate Std. Error t value Pr(>|t|)   
## (Intercept) 7.225e+00 6.544e+00 1.104 0.27831   
## poverty 4.117e-01 2.261e-01 1.821 0.07861 .   
## RMHI -2.948e-04 9.288e-05 -3.174 0.00347 \*\*  
## HPI -9.595e-03 7.008e-03 -1.369 0.18112   
## population 4.180e-06 2.116e-06 1.975 0.05752 .   
## sp500 -3.775e-02 1.700e-02 -2.221 0.03408 \*   
## ---  
## Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1  
##   
## Residual standard error: 1.446 on 30 degrees of freedom  
## Multiple R-squared: 0.5014, Adjusted R-squared: 0.4183   
## F-statistic: 6.033 on 5 and 30 DF, p-value: 0.0005614

anova(OR\_reg1)

## Analysis of Variance Table  
##   
## Response: value  
## Df Sum Sq Mean Sq F value Pr(>F)   
## poverty 1 25.221 25.2210 12.0612 0.001587 \*\*  
## RMHI 1 12.891 12.8909 6.1647 0.018850 \*   
## HPI 1 5.298 5.2980 2.5336 0.121929   
## population 1 9.355 9.3553 4.4739 0.042832 \*   
## sp500 1 10.311 10.3112 4.9310 0.034076 \*   
## Residuals 30 62.733 2.0911   
## ---  
## Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1

OR\_reg2 <- lm(value ~ HPI + poverty + sp500 + log\_pop + log\_RMHI, data = OR\_data)  
summary(OR\_reg2)

##   
## Call:  
## lm(formula = value ~ HPI + poverty + sp500 + log\_pop + log\_RMHI,   
## data = OR\_data)  
##   
## Residuals:  
## Min 1Q Median 3Q Max   
## -3.4042 -0.7362 -0.0734 0.6893 3.7150   
##   
## Coefficients:  
## Estimate Std. Error t value Pr(>|t|)   
## (Intercept) 1.207651 97.805741 0.012 0.99023   
## HPI -0.008184 0.006390 -1.281 0.21009   
## poverty 0.437489 0.224488 1.949 0.06072 .   
## sp500 -0.037701 0.017118 -2.202 0.03546 \*   
## log\_pop 12.531654 6.508638 1.925 0.06371 .   
## log\_RMHI -16.954334 5.392037 -3.144 0.00374 \*\*  
## ---  
## Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1  
##   
## Residual standard error: 1.456 on 30 degrees of freedom  
## Multiple R-squared: 0.4947, Adjusted R-squared: 0.4104   
## F-statistic: 5.873 on 5 and 30 DF, p-value: 0.0006737

anova(OR\_reg2)

## Analysis of Variance Table  
##   
## Response: value  
## Df Sum Sq Mean Sq F value Pr(>F)   
## HPI 1 0.628 0.6282 0.2964 0.590165   
## poverty 1 25.858 25.8582 12.2020 0.001504 \*\*  
## sp500 1 12.072 12.0718 5.6964 0.023507 \*   
## log\_pop 1 2.724 2.7236 1.2852 0.265914   
## log\_RMHI 1 20.952 20.9519 9.8868 0.003737 \*\*  
## Residuals 30 63.575 2.1192   
## ---  
## Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1

OR\_linearModelSignificant <- lm(value ~ poverty + log\_pop, data = OR\_data)  
summary(OR\_linearModelSignificant)

##   
## Call:  
## lm(formula = value ~ poverty + log\_pop, data = OR\_data)  
##   
## Residuals:  
## Min 1Q Median 3Q Max   
## -2.568 -1.150 -0.401 1.372 3.736   
##   
## Coefficients:  
## Estimate Std. Error t value Pr(>|t|)   
## (Intercept) 9.0377 31.3508 0.288 0.77494   
## poverty 0.6193 0.2142 2.891 0.00675 \*\*  
## log\_pop -0.6761 2.1097 -0.320 0.75063   
## ---  
## Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1  
##   
## Residual standard error: 1.743 on 33 degrees of freedom  
## Multiple R-squared: 0.203, Adjusted R-squared: 0.1546   
## F-statistic: 4.201 on 2 and 33 DF, p-value: 0.02369

CO\_data <- merged\_final[ merged\_final$state == "CO", ]

#Descriptive statistics  
stat.desc(CO\_data)

## year period periodName value state quarter annual  
## nbr.val 3.600000e+01 NA NA 36.0000000 NA 36 36  
## nbr.null 0.000000e+00 NA NA 0.0000000 NA 0 0  
## nbr.na 0.000000e+00 NA NA 0.0000000 NA 0 0  
## min 1.984000e+03 NA NA 2.5000000 NA 4 1  
## max 2.018000e+03 NA NA 8.8000000 NA 4 1  
## range 3.400000e+01 NA NA 6.3000000 NA 0 0  
## sum 7.204500e+04 NA NA 184.0000000 NA 144 36  
## median 2.001500e+03 NA NA 5.1500000 NA 4 1  
## mean 2.001250e+03 NA NA 5.1111111 NA 4 1  
## SE.mean 1.701715e+00 NA NA 0.3036753 NA 0 0  
## CI.mean.0.95 3.454665e+00 NA NA 0.6164937 NA 0 0  
## var 1.042500e+02 NA NA 3.3198730 NA 0 0  
## std.dev 1.021029e+01 NA NA 1.8220519 NA 0 0  
## coef.var 5.101956e-03 NA NA 0.3564884 NA 0 0  
## HPI MHI RMHI poverty population  
## nbr.val 3.600000e+01 3.600000e+01 3.600000e+01 36.0000000 3.600000e+01  
## nbr.null 0.000000e+00 0.000000e+00 0.000000e+00 0.0000000 0.000000e+00  
## nbr.na 0.000000e+00 0.000000e+00 0.000000e+00 0.0000000 0.000000e+00  
## min 1.235900e+02 2.580100e+04 5.256000e+04 8.2000000 3.169992e+06  
## max 5.702700e+02 7.498400e+04 7.681200e+04 13.7000000 5.695564e+06  
## range 4.466800e+02 4.918300e+04 2.425200e+04 5.5000000 2.525572e+06  
## sum 1.007002e+04 1.728672e+06 2.373177e+06 380.1000000 1.567548e+08  
## median 3.087500e+02 4.884550e+04 6.720500e+04 10.1000000 4.458047e+06  
## mean 2.797228e+02 4.801867e+04 6.592158e+04 10.5583333 4.354300e+06  
## SE.mean 2.059406e+01 2.469593e+03 1.052233e+03 0.2632934 1.355787e+05  
## CI.mean.0.95 4.180816e+01 5.013540e+03 2.136146e+03 0.5345140 2.752394e+05  
## var 1.526815e+04 2.195600e+08 3.985897e+07 2.4956429 6.617369e+11  
## std.dev 1.235644e+02 1.481756e+04 6.313396e+03 1.5797604 8.134721e+05  
## coef.var 4.417386e-01 3.085791e-01 9.577131e-02 0.1496221 1.868204e-01  
## sp500 log\_pop log\_RMHI  
## nbr.val 36.000000 36.00000000 3.600000e+01  
## nbr.null 1.000000 0.00000000 0.000000e+00  
## nbr.na 0.000000 0.00000000 0.000000e+00  
## min -38.490000 14.96923962 1.086971e+01  
## max 34.110000 15.55519818 1.124912e+01  
## range 72.600000 0.58595856 3.794055e-01  
## sum 341.220000 549.68892209 3.992972e+02  
## median 11.895000 15.31019488 1.111549e+01  
## mean 9.478333 15.26913672 1.109159e+01  
## SE.mean 2.652485 0.03193540 1.643015e-02  
## CI.mean.0.95 5.384831 0.06483231 3.335498e-02  
## var 253.284340 0.03671531 9.718198e-03  
## std.dev 15.914909 0.19161240 9.858092e-02  
## coef.var 1.679083 0.01254900 8.887898e-03

The dependent variable, unemployment, has a mean of 5.11 and a standard deviation of 1.82. The lowest unemployment rate was in 2000 at 2.5, while the highest was in 2010 at 8.8. The mean HPI was 279.72, with a standard deviation of 123.56. The lowest HPI was in 1988 at 123.59, while the highest was in 2018 at 570.27. Poverty rate had a mean of 10.56 and a standard deviation of 1.58. In 1997, lowest rate was seen of 8.2, and the highest was 13.7 in 1990. The sp500 index mean was 9.48, and the standard deviation was 15.70. In 2008, the lowest index was seen at -38.4, while the highest index was seen in 1995 at 34.11. The logged population mean was 15.30, and the standard deviation was 0.19. The lowest logged population at 14.97 in 1984, and the highest in 2018 at 15.56. Logged real median household income had a mean of 11.09 and a standard deviation of 0.099. In 1984, Missouri had the lowest log\_RMHI at 10.87, while had the highest in 2000 at 11.25.

#pairwise correlations  
sapply(CO\_data, class)

## year period periodName value state quarter   
## "numeric" "character" "character" "numeric" "character" "numeric"   
## annual HPI MHI RMHI poverty population   
## "numeric" "numeric" "numeric" "numeric" "numeric" "numeric"   
## sp500 log\_pop log\_RMHI   
## "numeric" "numeric" "numeric"

sapply(CO\_data, is.factor)

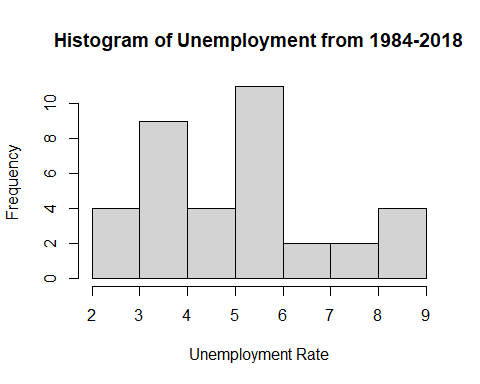
## year period periodName value state quarter annual   
## FALSE FALSE FALSE FALSE FALSE FALSE FALSE   
## HPI MHI RMHI poverty population sp500 log\_pop   
## FALSE FALSE FALSE FALSE FALSE FALSE FALSE   
## log\_RMHI   
## FALSE

cor(CO\_data[sapply(CO\_data, function(x) !is.character(x))])

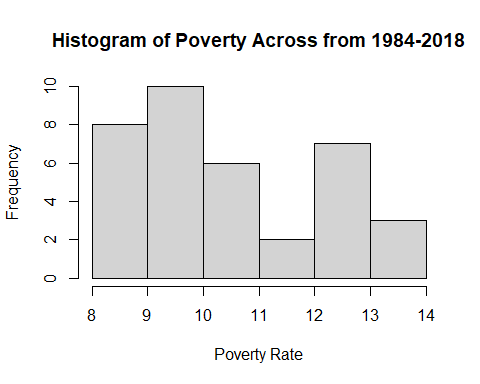
## Warning in cor(CO\_data[sapply(CO\_data, function(x) !is.character(x))]): the  
## standard deviation is zero

## year value quarter annual HPI MHI  
## year 1.00000000 -0.09998024 NA NA 0.9595940 0.9832409  
## value -0.09998024 1.00000000 NA NA -0.2161680 -0.1719434  
## quarter NA NA 1 NA NA NA  
## annual NA NA NA 1 NA NA  
## HPI 0.95959395 -0.21616797 NA NA 1.0000000 0.9670726  
## MHI 0.98324088 -0.17194344 NA NA 0.9670726 1.0000000  
## RMHI 0.74689505 -0.37443856 NA NA 0.7891121 0.8446744  
## poverty -0.10366786 0.72358186 NA NA -0.2306553 -0.2101074  
## population 0.99615532 -0.09548704 NA NA 0.9674495 0.9866006  
## sp500 -0.13852928 0.01642260 NA NA -0.1966361 -0.1404192  
## log\_pop 0.99343100 -0.10580472 NA NA 0.9591200 0.9850953  
## log\_RMHI 0.74518017 -0.36219627 NA NA 0.7838381 0.8406593  
## RMHI poverty population sp500 log\_pop  
## year 0.74689505 -0.1036679 0.99615532 -0.13852928 0.9934310  
## value -0.37443856 0.7235819 -0.09548704 0.01642260 -0.1058047  
## quarter NA NA NA NA NA  
## annual NA NA NA NA NA  
## HPI 0.78911214 -0.2306553 0.96744947 -0.19663610 0.9591200  
## MHI 0.84467441 -0.2101074 0.98660060 -0.14041918 0.9850953  
## RMHI 1.00000000 -0.5253135 0.77453905 -0.09012611 0.7927908  
## poverty -0.52531353 1.0000000 -0.12836813 -0.06261130 -0.1480198  
## population 0.77453905 -0.1283681 1.00000000 -0.14975993 0.9971220  
## sp500 -0.09012611 -0.0626113 -0.14975993 1.00000000 -0.1525337  
## log\_pop 0.79279075 -0.1480198 0.99712200 -0.15253373 1.0000000  
## log\_RMHI 0.99875549 -0.5221603 0.77381784 -0.09397732 0.7936630  
## log\_RMHI  
## year 0.74518017  
## value -0.36219627  
## quarter NA  
## annual NA  
## HPI 0.78383810  
## MHI 0.84065927  
## RMHI 0.99875549  
## poverty -0.52216030  
## population 0.77381784  
## sp500 -0.09397732  
## log\_pop 0.79366297  
## log\_RMHI 1.00000000

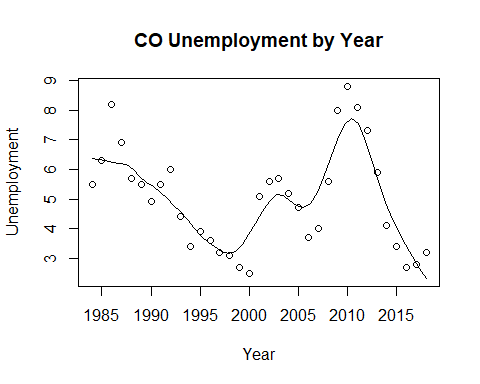
hist(CO\_data$value,  
 main="Histogram of Unemployment from 1984-2018",  
 xlab="Unemployment Rate",  
 breaks = 5)

 In Colorado, from 1984 to 2018, the annual unemployment rates ranged from 2.0% to 8.9%. 11 of the 36 years were characterized by unemployment rates between the range of 5.0-5.9%.

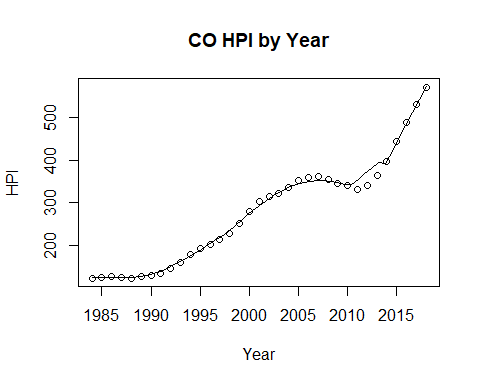
hist(CO\_data$poverty,  
 main="Histogram of Poverty Across from 1984-2018",  
 xlab="Poverty Rate",  
 breaks = 5)

 In Colorado, from 1984 to 2018, the poverty rates ranged from 8.0% to 13.9%. The histogram displays a slightly positively skewed distribution. 24 of the 36 years were characterized by poverty rates in the three ranges with the lowest values, 8.0-8.9%, 9.0-9.9%, and 10.0-10.9%.

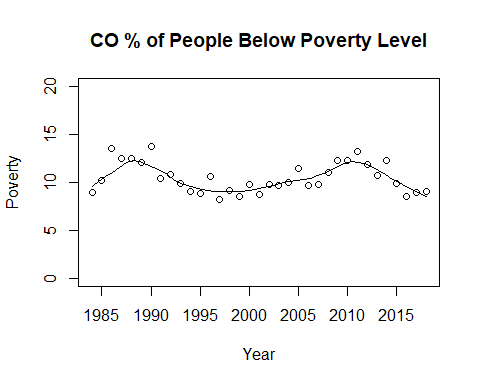
scatter.smooth(CO\_data$year, CO\_data$value, main="CO Unemployment by Year", xlab = "Year", ylab = "Unemployment", span = 2/8)



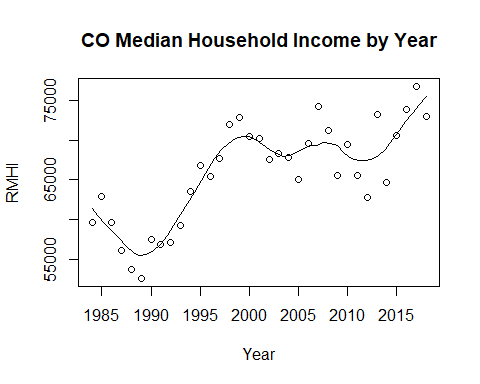
scatter.smooth(CO\_data$year, CO\_data$HPI, main="CO HPI by Year", xlab = "Year", ylab = "HPI", span = 2/8)



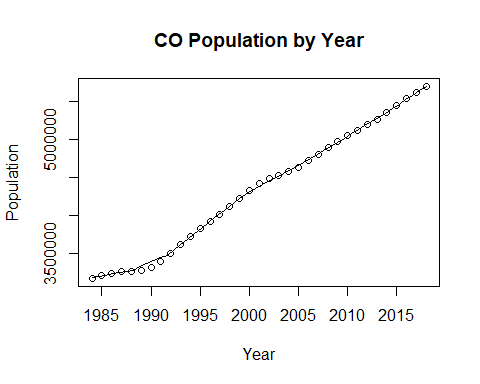
scatter.smooth(x=CO\_data$year, y=CO\_data$poverty, main="CO % of People Below Poverty Level", xlab = "Year", ylab = "Poverty", ylim = c(0,20), span = 2/8)



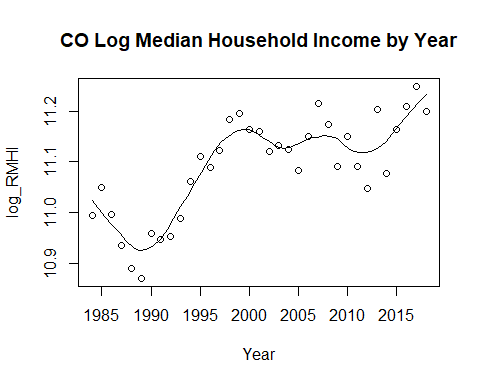
CO\_data$RMHI <- as.integer(CO\_data$RMHI)  
scatter.smooth(x=CO\_data$year, y=CO\_data$RMHI, main="CO Median Household Income by Year", xlab = "Year", ylab = "RMHI", span = 2/8 )



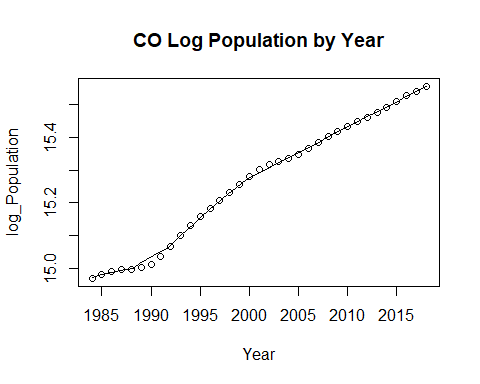
CO\_data$population <- as.integer(CO\_data$population)  
scatter.smooth(x=CO\_data$year, y=CO\_data$population, main="CO Population by Year", xlab = "Year", ylab = "Population", span = 2/8 )



#CO\_data$log\_RMHI <- as.integer(CO\_data$log\_RMHI)  
scatter.smooth(x=CO\_data$year, y=CO\_data$log\_RMHI, main="CO Log Median Household Income by Year", xlab = "Year", ylab = "log\_RMHI", span = 2/8 )



#CO\_data$log\_pop <- as.integer(CO\_data$log\_pop)  
scatter.smooth(x=CO\_data$year, y=CO\_data$log\_pop, main="CO Log Population by Year", xlab = "Year", ylab = "log\_Population", span = 2/8)



cor(CO\_data$RMHI, CO\_data$value)

## [1] -0.3744386

cor(CO\_data$HPI, CO\_data$value)

## [1] -0.216168

cor(CO\_data$poverty, CO\_data$value)

## [1] 0.7235819

cor(CO\_data$population, CO\_data$value)

## [1] -0.09548704

cor(CO\_data$sp500, CO\_data$value)

## [1] 0.0164226

cor(CO\_data$log\_pop, CO\_data$value)

## [1] -0.1058047

cor(CO\_data$log\_RMHI, CO\_data$value)

## [1] -0.3621963

#run base correlations between categories and dependent variable

The correlation between the median household income and unemployment in Colorado is -0.3744386. The correlation between the median household income and unemployment is a medium negative correlation. The correlation suggests an inverse relationship between the two variables.

The correlation between the housing price index and unemployment in Colorado is -0.216168. The correlation between the housing price index and unemployment shows a weak negative correlation between the variables. The negative correlation illustrates the possibility of there being an inverse relationship between the two variables. The closeness of the correlation to 0 does show that there is a possibility that maybe the variables do not have any correlation at all. Of course with all of the depressions that have occurred and the pandemic this year, this could skew the correlation and downplay the relationship between the two variables.

The correlation between poverty, the proportion of the population below the poverty line, and unemployment in Colorado is 0.7235819. The correlation between poverty and unemployment is a large positive correlation. A strong positive correlation suggests that there is a relationship between the two variables and it suggests that the two variables tend to move in the same direction of one another.

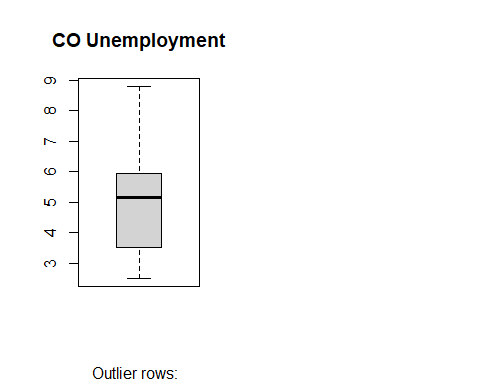
The correlation between the population and unemployment in Colorado is -0.09548704. The correlation between the population and unemployment is a small negative correlation. The relationship between the two variables is quite small due to the correlation being so close to 0.

THe correlation between the S&P 500 and unemployment is 0.0164226. The correlation between the S%P 500 and unemployment is an extremely small correlation. With the correlation being so close to 0, it signifies that the relationship between the two variables is not that common. Likewise, the correlation shows that there is a chance that there is no direct relationship between the two variables.

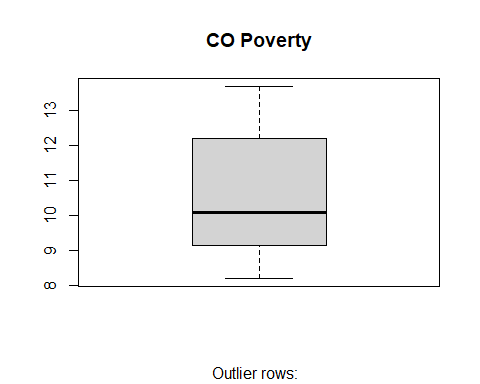
The correlation between the log of the population and unemployment in Colorado is -0.1058047. The correlation between the log of the population and unemployment is a weak negative correlation. Just like in the case of the S&P 500, the weak correlation does not show a strong inverse relationship between the two variables.

The correlation between the log of the median household income and unemployment in Colorado is -0.3621963. THe correlation between the log of the median household income and unemployment is that of a medium negative correlation. The correlation between these two variables in Colorado is much stronger than the S&P 500 and the log of the median household income. With the negative relationship, it suggests that the two variables tend to move in an inverse fashion.

par(mfrow=c(1, 2)) # divide graph area in 2 columns  
boxplot(CO\_data$value, main="CO Unemployment", sub=paste("Outlier rows: ", boxplot.stats(CO\_data$value)$out)) # box plot for 'Unemployment'



boxplot(CO\_data$poverty, main="CO Poverty", sub=paste("Outlier rows: ", boxplot.stats(CO\_data$poverty)$out)) # box plot for 'Poverty'



#create box plot for poverty level and unemployment level

#run multiple linear model for data  
CO\_reg1 <- lm(value ~ poverty + RMHI + HPI + population + sp500, data = CO\_data)  
summary(CO\_reg1)

##   
## Call:  
## lm(formula = value ~ poverty + RMHI + HPI + population + sp500,   
## data = CO\_data)  
##   
## Residuals:  
## Min 1Q Median 3Q Max   
## -2.61909 -0.68601 -0.02734 0.87070 1.89876   
##   
## Coefficients:  
## Estimate Std. Error t value Pr(>|t|)   
## (Intercept) -7.261e+00 5.222e+00 -1.390 0.1746   
## poverty 7.298e-01 2.074e-01 3.519 0.0014 \*\*  
## RMHI 2.755e-06 7.499e-05 0.037 0.9709   
## HPI -1.298e-02 8.059e-03 -1.611 0.1177   
## population 1.862e-06 1.284e-06 1.450 0.1575   
## sp500 9.496e-04 1.453e-02 0.065 0.9483   
## ---  
## Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1  
##   
## Residual standard error: 1.298 on 30 degrees of freedom  
## Multiple R-squared: 0.5653, Adjusted R-squared: 0.4929   
## F-statistic: 7.804 on 5 and 30 DF, p-value: 8.396e-05

anova(CO\_reg1)

## Analysis of Variance Table  
##   
## Response: value  
## Df Sum Sq Mean Sq F value Pr(>F)   
## poverty 1 60.837 60.837 36.1367 1.351e-06 \*\*\*  
## RMHI 1 0.005 0.005 0.0031 0.9562   
## HPI 1 1.044 1.044 0.6201 0.4372   
## population 1 3.797 3.797 2.2556 0.1436   
## sp500 1 0.007 0.007 0.0043 0.9483   
## Residuals 30 50.505 1.684   
## ---  
## Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1

CO\_reg2 <- lm(value ~ HPI + poverty + sp500 + log\_pop + log\_RMHI, data = CO\_data)  
summary(CO\_reg2)

##   
## Call:  
## lm(formula = value ~ HPI + poverty + sp500 + log\_pop + log\_RMHI,   
## data = CO\_data)  
##   
## Residuals:  
## Min 1Q Median 3Q Max   
## -2.53081 -0.70979 0.02929 0.74503 1.98297   
##   
## Coefficients:  
## Estimate Std. Error t value Pr(>|t|)   
## (Intercept) -96.635103 66.415495 -1.455 0.156052   
## HPI -0.009979 0.006876 -1.451 0.157043   
## poverty 0.778968 0.203094 3.836 0.000599 \*\*\*  
## sp500 0.002700 0.014457 0.187 0.853084   
## log\_pop 5.951883 4.909061 1.212 0.234811   
## log\_RMHI 0.487516 5.013722 0.097 0.923185   
## ---  
## Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1  
##   
## Residual standard error: 1.306 on 30 degrees of freedom  
## Multiple R-squared: 0.5596, Adjusted R-squared: 0.4862   
## F-statistic: 7.624 on 5 and 30 DF, p-value: 0.0001009

anova(CO\_reg2)

## Analysis of Variance Table  
##   
## Response: value  
## Df Sum Sq Mean Sq F value Pr(>F)   
## HPI 1 5.430 5.430 3.1831 0.08452 .   
## poverty 1 55.705 55.705 32.6568 3.103e-06 \*\*\*  
## sp500 1 0.315 0.315 0.1848 0.67034   
## log\_pop 1 3.557 3.557 2.0851 0.15911   
## log\_RMHI 1 0.016 0.016 0.0095 0.92319   
## Residuals 30 51.173 1.706   
## ---  
## Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1

CO\_linearModelSignificant <- lm(value ~ poverty + log\_pop, data = CO\_data)  
summary(CO\_linearModelSignificant)

##   
## Call:  
## lm(formula = value ~ poverty + log\_pop, data = CO\_data)  
##   
## Residuals:  
## Min 1Q Median 3Q Max   
## -2.83048 -0.88120 0.08845 0.85265 2.23288   
##   
## Coefficients:  
## Estimate Std. Error t value Pr(>|t|)   
## (Intercept) -3.89578 17.92034 -0.217 0.829   
## poverty 0.83479 0.14013 5.957 1.09e-06 \*\*\*  
## log\_pop 0.01264 1.15529 0.011 0.991   
## ---  
## Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1  
##   
## Residual standard error: 1.295 on 33 degrees of freedom  
## Multiple R-squared: 0.5236, Adjusted R-squared: 0.4947   
## F-statistic: 18.13 on 2 and 33 DF, p-value: 4.863e-06

WV\_data <- merged\_final[ merged\_final$state == "WV", ]

#Descriptive statistics  
stat.desc(WV\_data)

## year period periodName value state quarter annual  
## nbr.val 3.600000e+01 NA NA 36.0000000 NA 36 36  
## nbr.null 0.000000e+00 NA NA 0.0000000 NA 0 0  
## nbr.na 0.000000e+00 NA NA 0.0000000 NA 0 0  
## min 1.984000e+03 NA NA 4.2000000 NA 4 1  
## max 2.018000e+03 NA NA 14.4000000 NA 4 1  
## range 3.400000e+01 NA NA 10.2000000 NA 0 0  
## sum 7.204500e+04 NA NA 260.6000000 NA 144 36  
## median 2.001500e+03 NA NA 6.8000000 NA 4 1  
## mean 2.001250e+03 NA NA 7.2388889 NA 4 1  
## SE.mean 1.701715e+00 NA NA 0.4132388 NA 0 0  
## CI.mean.0.95 3.454665e+00 NA NA 0.8389194 NA 0 0  
## var 1.042500e+02 NA NA 6.1475873 NA 0 0  
## std.dev 1.021029e+01 NA NA 2.4794329 NA 0 0  
## coef.var 5.101956e-03 NA NA 0.3425157 NA 0 0  
## HPI MHI RMHI poverty population  
## nbr.val 36.0000000 3.600000e+01 3.600000e+01 36.0000000 3.600000e+01  
## nbr.null 0.0000000 0.000000e+00 0.000000e+00 0.0000000 0.000000e+00  
## nbr.na 0.0000000 0.000000e+00 0.000000e+00 0.0000000 0.000000e+00  
## min 87.2800000 1.598300e+04 3.566200e+04 14.2000000 1.792548e+06  
## max 230.8600000 5.057300e+04 5.110100e+04 22.4000000 1.927697e+06  
## range 143.5800000 3.459000e+04 1.543900e+04 8.2000000 1.351490e+05  
## sum 5907.9000000 1.140958e+06 1.561645e+06 633.5000000 6.597333e+07  
## median 164.2650000 2.954200e+04 4.301200e+04 16.9500000 1.823254e+06  
## mean 164.1083333 3.169328e+04 4.337903e+04 17.5972222 1.832592e+06  
## SE.mean 8.1120264 1.692733e+03 7.148259e+02 0.4040762 4.925293e+03  
## CI.mean.0.95 16.4682892 3.436430e+03 1.451174e+03 0.8203183 9.998876e+03  
## var 2368.9790257 1.031524e+08 1.839514e+07 5.8779921 8.733063e+08  
## std.dev 48.6721586 1.015640e+04 4.288955e+03 2.4244571 2.955176e+04  
## coef.var 0.2965855 3.204590e-01 9.887163e-02 0.1377750 1.612566e-02  
## sp500 log\_pop log\_RMHI  
## nbr.val 36.000000 3.600000e+01 3.600000e+01  
## nbr.null 1.000000 0.000000e+00 0.000000e+00  
## nbr.na 0.000000 0.000000e+00 0.000000e+00  
## min -38.490000 1.439915e+01 1.048184e+01  
## max 34.110000 1.447184e+01 1.084156e+01  
## range 72.600000 7.268795e-02 3.597184e-01  
## sum 341.220000 5.191602e+02 3.842244e+02  
## median 11.895000 1.441613e+01 1.066923e+01  
## mean 9.478333 1.442112e+01 1.067290e+01  
## SE.mean 2.652485 2.659874e-03 1.669191e-02  
## CI.mean.0.95 5.384831 5.399832e-03 3.388637e-02  
## var 253.284340 2.546975e-04 1.003031e-02  
## std.dev 15.914909 1.595925e-02 1.001514e-01  
## coef.var 1.679083 1.106658e-03 9.383714e-03

The dependent variable, unemployment, has a mean of 7.25 and a standard deviation of 2.48. The lowest unemployment rate was in 2006 at 4.2 while the highest was in 1984 at 14.40. The mean HPI was 164.11, with a standard deviation of 48.67. The lowest HPI was in 1984 at 87.28, while the highest was in 2018 at 230.86. Poverty rate had a mean of 17.60 and a standard deviation of 2.42. In 2004, lowest rate was seen of 14.2, and the highest was 22.4 in 1986. The sp500 index mean was 9.48, and the standard deviation was 15.70. In 2008, the lowest index was seen at -38.4, while the highest index was seen in 1995 at 34.11. The logged population mean was 14.42, and the standard deviation was 0.016. The lowest logged population at 14.40 in 1990, and the highest in 1984 at 14.47. Logged real median household income had a mean of 10.67 and a standard deviation of 0.10. In 1992, West Virginia had the lowest log\_RMHI at 10.48, while had the highest in 2007 at 10.84.

#pairwise correlations  
sapply(WV\_data, class)

## year period periodName value state quarter   
## "numeric" "character" "character" "numeric" "character" "numeric"   
## annual HPI MHI RMHI poverty population   
## "numeric" "numeric" "numeric" "numeric" "numeric" "numeric"   
## sp500 log\_pop log\_RMHI   
## "numeric" "numeric" "numeric"

sapply(WV\_data, is.factor)

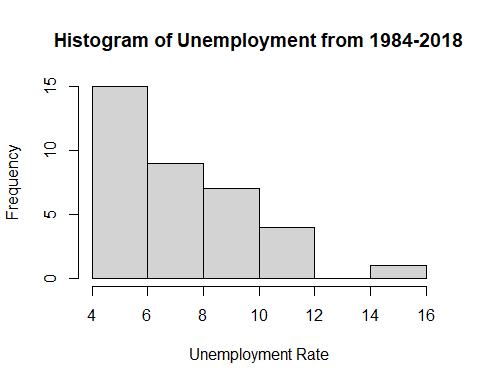
## year period periodName value state quarter annual   
## FALSE FALSE FALSE FALSE FALSE FALSE FALSE   
## HPI MHI RMHI poverty population sp500 log\_pop   
## FALSE FALSE FALSE FALSE FALSE FALSE FALSE   
## log\_RMHI   
## FALSE

cor(WV\_data[sapply(WV\_data, function(x) !is.character(x))])

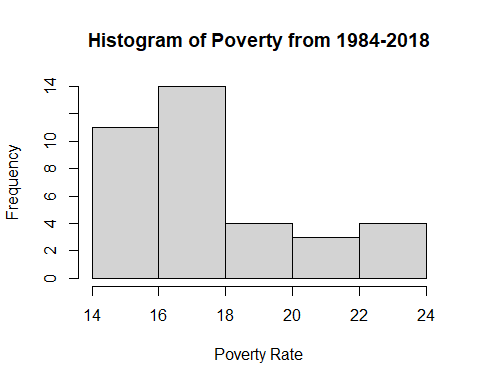
## Warning in cor(WV\_data[sapply(WV\_data, function(x) !is.character(x))]): the  
## standard deviation is zero

## year value quarter annual HPI MHI  
## year 1.0000000 -0.6830875 NA NA 0.97971522 0.98286504  
## value -0.6830875 1.0000000 NA NA -0.71386948 -0.64676220  
## quarter NA NA 1 NA NA NA  
## annual NA NA NA 1 NA NA  
## HPI 0.9797152 -0.7138695 NA NA 1.00000000 0.97903651  
## MHI 0.9828650 -0.6467622 NA NA 0.97903651 1.00000000  
## RMHI 0.8376848 -0.6170457 NA NA 0.87205819 0.91397010  
## poverty -0.5003150 0.6996082 NA NA -0.55206136 -0.54467318  
## population -0.1232569 0.5144306 NA NA -0.08724377 -0.08789167  
## sp500 -0.1385293 0.1979507 NA NA -0.16786502 -0.11281889  
## log\_pop -0.1170711 0.5093770 NA NA -0.08085322 -0.08195948  
## log\_RMHI 0.8388360 -0.6322822 NA NA 0.87065335 0.91113096  
## RMHI poverty population sp500 log\_pop  
## year 0.83768478 -0.5003150 -0.12325691 -0.13852928 -0.11707114  
## value -0.61704566 0.6996082 0.51443058 0.19795068 0.50937703  
## quarter NA NA NA NA NA  
## annual NA NA NA NA NA  
## HPI 0.87205819 -0.5520614 -0.08724377 -0.16786502 -0.08085322  
## MHI 0.91397010 -0.5446732 -0.08789167 -0.11281889 -0.08195948  
## RMHI 1.00000000 -0.7093226 -0.14833523 -0.02144328 -0.14370564  
## poverty -0.70932258 1.0000000 0.42128203 0.19613672 0.41901962  
## population -0.14833523 0.4212820 1.00000000 0.09952714 0.99993588  
## sp500 -0.02144328 0.1961367 0.09952714 1.00000000 0.10072129  
## log\_pop -0.14370564 0.4190196 0.99993588 0.10072129 1.00000000  
## log\_RMHI 0.99860531 -0.7294518 -0.16973197 -0.02206507 -0.16515363  
## log\_RMHI  
## year 0.83883601  
## value -0.63228217  
## quarter NA  
## annual NA  
## HPI 0.87065335  
## MHI 0.91113096  
## RMHI 0.99860531  
## poverty -0.72945180  
## population -0.16973197  
## sp500 -0.02206507  
## log\_pop -0.16515363  
## log\_RMHI 1.00000000

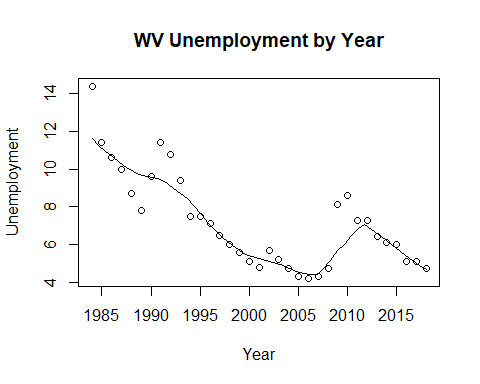
hist(WV\_data$value,  
 main="Histogram of Unemployment from 1984-2018",  
 xlab="Unemployment Rate",  
 breaks = 5)

 In West Virginia,from 1984 to 2018, the unemployment rates ranged from 4.0-15.9%. The histogram displays a positively skewed distribution. 15 of the 36 years were characterized by unemployment rates ranging in lowest interval, 4.0-5.9%.

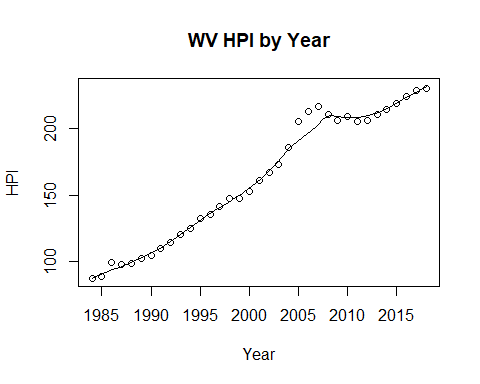
hist(WV\_data$poverty,  
 main="Histogram of Poverty from 1984-2018",  
 xlab="Poverty Rate",  
 breaks = 5)

 In West Virginia, from 1984 to 2018, the annual poverty rates ranged from 14.0% to 23.9%. This is the highest range out of all 8 states examined. Although this range has larger values comparatively, the histogram displays a positively skewed distribution. 25 out of the 36 years were characterized by unemployment rates that fell into the first two intervals, 14.0-15.9% & 16-17.9%.

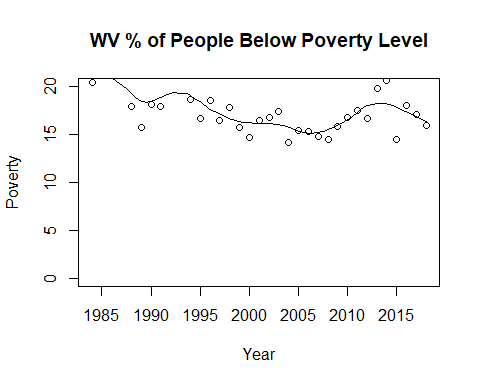
scatter.smooth(WV\_data$year, WV\_data$value, main="WV Unemployment by Year", xlab = "Year", ylab = "Unemployment", span = 2/8)



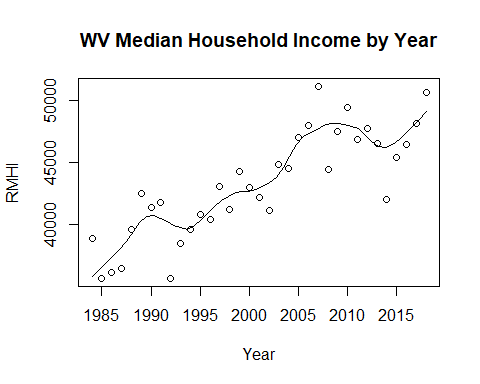
scatter.smooth(WV\_data$year, WV\_data$HPI, main="WV HPI by Year", xlab = "Year", ylab = "HPI", span = 2/8)



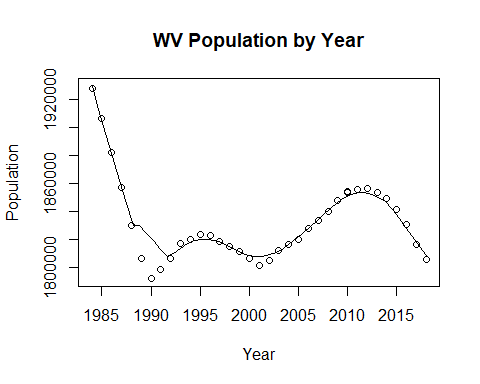
scatter.smooth(x=WV\_data$year, y=WV\_data$poverty, main="WV % of People Below Poverty Level", xlab = "Year", ylab = "Poverty", ylim = c(0,20), span = 2/8)



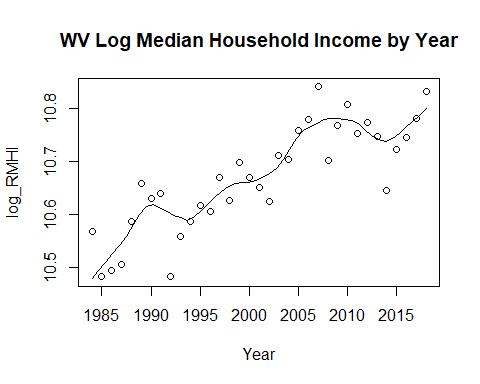
WV\_data$RMHI <- as.integer(WV\_data$RMHI)  
scatter.smooth(x=WV\_data$year, y=WV\_data$RMHI, main="WV Median Household Income by Year", xlab = "Year", ylab = "RMHI", span = 2/8 )



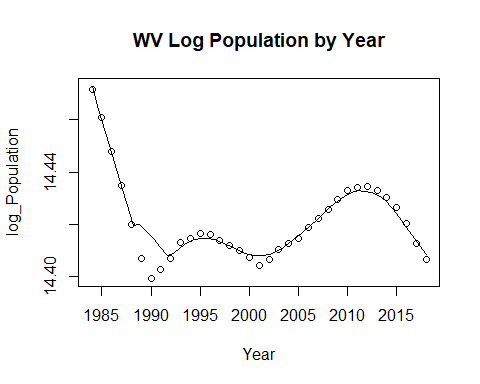
WV\_data$population <- as.integer(WV\_data$population)  
scatter.smooth(x=WV\_data$year, y=WV\_data$population, main="WV Population by Year", xlab = "Year", ylab = "Population", span = 2/8 )



#WV\_data$log\_RMHI <- as.integer(WV\_data$log\_RMHI)  
scatter.smooth(x=WV\_data$year, y=WV\_data$log\_RMHI, main="WV Log Median Household Income by Year", xlab = "Year", ylab = "log\_RMHI", span = 2/8 )



#WV\_data$log\_pop <- as.integer(WV\_data$log\_pop)  
scatter.smooth(x=WV\_data$year, y=WV\_data$log\_pop, main="WV Log Population by Year", xlab = "Year", ylab = "log\_Population", span = 2/8)



cor(WV\_data$RMHI, WV\_data$value)

## [1] -0.6170457

cor(WV\_data$HPI, WV\_data$value)

## [1] -0.7138695

cor(WV\_data$poverty, WV\_data$value)

## [1] 0.6996082

cor(WV\_data$population, WV\_data$value)

## [1] 0.5144306

cor(WV\_data$sp500, WV\_data$value)

## [1] 0.1979507

cor(WV\_data$log\_pop, WV\_data$value)

## [1] 0.509377

cor(WV\_data$log\_RMHI, WV\_data$value)

## [1] -0.6322822

#run base correlations between categories and dependent variable

The correlation between the median household income and unemployment in West Virginia is -0.6170457. The correlation between the median household income and unemployment is that of a large negative correlation. A large negative correlation suggests that in most cases the variables tend to move in inverse directions.

The correlation between the housing price index and unemployment in West Virginia is -0.7138695. The correlation between the housing price index and unemployment is that of a large negative correlation. Just like with the median household income, the housing price index suggests an inverse relationship between unemployment.

The correlation between poverty, the proportion of the population below the poverty line, and unemployment in West Virginia is 0.6996082. The correlation between poverty and unemployment is that of a large positive correlation. A large positive correlation suggests that the two variables tend to move in the same direction unlike the median household income and the housing price index.

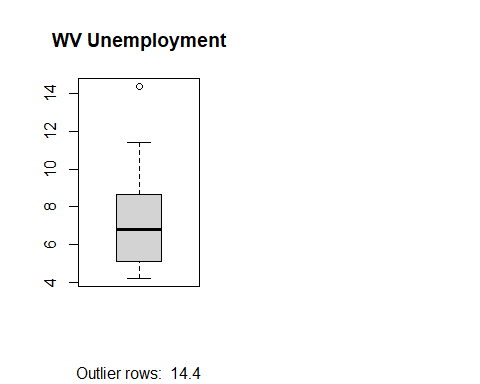
The correlation between the population and unemployment is 0.5144306. The correlation between the population and unemployment is that of a large positive correlation. Just like with poverty, the correlation suggests that the two variables tend to move in the same direction of one another.

The correlation between the S&P 500 and unemployment in West Virginia is 0.1979507. The correlation between the S&P 500 and unemployment is a small positive correlation. The correlation shows that there is a small positive relationship between the two variables signifying that the two variables tend to move in the same direction.

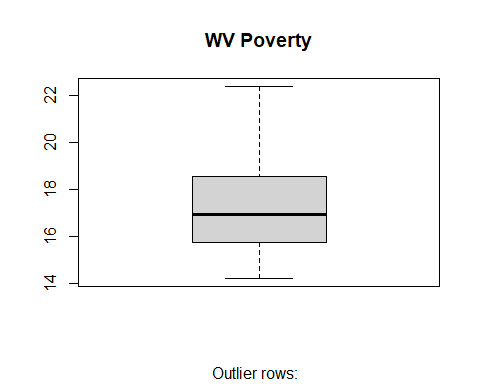
The correlation between the log of the population and unemployment in West Virginia is 0.509377. The correlation between the log of the population and unemployment is that of a large positive Correlation. The positive correlation suggests that the two variables mimic one another.

The correlation between the log of the median household income and unemployment in West Virginia is -0.6322822. THe correlation between the log of the median household income and unemployment is a large negative correlation. The large negative correlation suggests that as the log of the median household income increases, the unemployment tends to decrease.

par(mfrow=c(1, 2)) # divide graph area in 2 columns  
boxplot(WV\_data$value, main="WV Unemployment", sub=paste("Outlier rows: ", boxplot.stats(WV\_data$value)$out)) # box plot for 'Unemployment'



boxplot(WV\_data$poverty, main="WV Poverty", sub=paste("Outlier rows: ", boxplot.stats(WV\_data$poverty)$out)) # box plot for 'Poverty'



#create box plot for poverty level and unemployment level

#run multiple linear model for data  
WV\_reg1 <- lm(value ~ poverty + RMHI + HPI + population + sp500, data = WV\_data)  
summary(WV\_reg1)

##   
## Call:  
## lm(formula = value ~ poverty + RMHI + HPI + population + sp500,   
## data = WV\_data)  
##   
## Residuals:  
## Min 1Q Median 3Q Max   
## -1.8385 -0.7966 -0.3535 0.4722 3.0570   
##   
## Coefficients:  
## Estimate Std. Error t value Pr(>|t|)   
## (Intercept) -5.897e+01 1.377e+01 -4.281 0.000176 \*\*\*  
## poverty 4.883e-01 1.418e-01 3.443 0.001719 \*\*   
## RMHI 3.567e-04 1.278e-04 2.792 0.009038 \*\*   
## HPI -4.956e-02 9.317e-03 -5.320 9.45e-06 \*\*\*  
## population 2.750e-05 7.756e-06 3.545 0.001310 \*\*   
## sp500 -1.221e-02 1.416e-02 -0.862 0.395348   
## ---  
## Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1  
##   
## Residual standard error: 1.195 on 30 degrees of freedom  
## Multiple R-squared: 0.8009, Adjusted R-squared: 0.7677   
## F-statistic: 24.14 on 5 and 30 DF, p-value: 1.105e-09

anova(WV\_reg1)

## Analysis of Variance Table  
##   
## Response: value  
## Df Sum Sq Mean Sq F value Pr(>F)   
## poverty 1 105.313 105.313 73.7581 1.396e-09 \*\*\*  
## RMHI 1 6.319 6.319 4.4257 0.043891 \*   
## HPI 1 41.472 41.472 29.0460 7.757e-06 \*\*\*  
## population 1 18.165 18.165 12.7220 0.001236 \*\*   
## sp500 1 1.062 1.062 0.7436 0.395348   
## Residuals 30 42.835 1.428   
## ---  
## Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1

WV\_reg2 <- lm(value ~ HPI + poverty + sp500 + log\_pop + log\_RMHI, data = WV\_data)  
summary(WV\_reg2)

##   
## Call:  
## lm(formula = value ~ HPI + poverty + sp500 + log\_pop + log\_RMHI,   
## data = WV\_data)  
##   
## Residuals:  
## Min 1Q Median 3Q Max   
## -1.6926 -0.7540 -0.3264 0.4607 3.0139   
##   
## Coefficients:  
## Estimate Std. Error t value Pr(>|t|)   
## (Intercept) -9.132e+02 2.079e+02 -4.393 0.000129 \*\*\*  
## HPI -5.035e-02 9.467e-03 -5.319 9.47e-06 \*\*\*  
## poverty 5.165e-01 1.477e-01 3.496 0.001492 \*\*   
## sp500 -1.345e-02 1.431e-02 -0.940 0.354765   
## log\_pop 5.187e+01 1.431e+01 3.624 0.001061 \*\*   
## log\_RMHI 1.609e+01 5.719e+00 2.814 0.008561 \*\*   
## ---  
## Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1  
##   
## Residual standard error: 1.195 on 30 degrees of freedom  
## Multiple R-squared: 0.801, Adjusted R-squared: 0.7678   
## F-statistic: 24.15 on 5 and 30 DF, p-value: 1.1e-09

anova(WV\_reg2)

## Analysis of Variance Table  
##   
## Response: value  
## Df Sum Sq Mean Sq F value Pr(>F)   
## HPI 1 109.650 109.650 76.8172 8.982e-10 \*\*\*  
## poverty 1 28.886 28.886 20.2367 9.559e-05 \*\*\*  
## sp500 1 0.240 0.240 0.1680 0.6847914   
## log\_pop 1 22.265 22.265 15.5984 0.0004386 \*\*\*  
## log\_RMHI 1 11.301 11.301 7.9171 0.0085606 \*\*   
## Residuals 30 42.823 1.427   
## ---  
## Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1

WV\_linearModelSignificant <- lm(value ~ poverty + log\_pop, data = WV\_data)  
summary(WV\_linearModelSignificant)

##   
## Call:  
## lm(formula = value ~ poverty + log\_pop, data = WV\_data)  
##   
## Residuals:  
## Min 1Q Median 3Q Max   
## -3.3284 -0.9962 -0.2468 1.0728 4.7333   
##   
## Coefficients:  
## Estimate Std. Error t value Pr(>|t|)   
## (Intercept) -590.9981 288.4061 -2.049 0.0485 \*   
## poverty 0.6031 0.1321 4.566 6.59e-05 \*\*\*  
## log\_pop 40.7475 20.0659 2.031 0.0504 .   
## ---  
## Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1  
##   
## Residual standard error: 1.72 on 33 degrees of freedom  
## Multiple R-squared: 0.5462, Adjusted R-squared: 0.5187   
## F-statistic: 19.86 on 2 and 33 DF, p-value: 2.182e-06

ME\_data <- merged\_final[ merged\_final$state == "ME", ]

#Descriptive statistics  
stat.desc(ME\_data)

## year period periodName value state quarter annual  
## nbr.val 3.600000e+01 NA NA 36.0000000 NA 36 36  
## nbr.null 0.000000e+00 NA NA 0.0000000 NA 0 0  
## nbr.na 0.000000e+00 NA NA 0.0000000 NA 0 0  
## min 1.984000e+03 NA NA 3.0000000 NA 4 1  
## max 2.018000e+03 NA NA 8.1000000 NA 4 1  
## range 3.400000e+01 NA NA 5.1000000 NA 0 0  
## sum 7.204500e+04 NA NA 189.7000000 NA 144 36  
## median 2.001500e+03 NA NA 4.8000000 NA 4 1  
## mean 2.001250e+03 NA NA 5.2694444 NA 4 1  
## SE.mean 1.701715e+00 NA NA 0.2623405 NA 0 0  
## CI.mean.0.95 3.454665e+00 NA NA 0.5325795 NA 0 0  
## var 1.042500e+02 NA NA 2.4776111 NA 0 0  
## std.dev 1.021029e+01 NA NA 1.5740429 NA 0 0  
## coef.var 5.101956e-03 NA NA 0.2987114 NA 0 0  
## HPI MHI RMHI poverty population  
## nbr.val 3.600000e+01 3.600000e+01 3.600000e+01 36.0000000 3.600000e+01  
## nbr.null 0.000000e+00 0.000000e+00 0.000000e+00 0.0000000 0.000000e+00  
## nbr.na 0.000000e+00 0.000000e+00 0.000000e+00 0.0000000 0.000000e+00  
## min 1.411200e+02 2.051900e+04 4.579300e+04 9.4000000 1.155635e+06  
## max 5.484600e+02 5.866300e+04 5.933800e+04 15.4000000 1.338404e+06  
## range 4.073400e+02 3.814400e+04 1.354500e+04 6.0000000 1.827690e+05  
## sum 1.267296e+04 1.397678e+06 1.933397e+06 429.9000000 4.597438e+07  
## median 3.451250e+02 3.718950e+04 5.455400e+04 11.8000000 1.290826e+06  
## mean 3.520267e+02 3.882439e+04 5.370547e+04 11.9416667 1.277066e+06  
## SE.mean 2.094329e+01 1.793819e+03 5.368914e+02 0.2341948 9.339756e+03  
## CI.mean.0.95 4.251713e+01 3.641645e+03 1.089948e+03 0.4754408 1.896071e+04  
## var 1.579037e+04 1.158403e+08 1.037709e+07 1.9745000 3.140317e+09  
## std.dev 1.256597e+02 1.076291e+04 3.221349e+03 1.4051690 5.603853e+04  
## coef.var 3.569608e-01 2.772204e-01 5.998176e-02 0.1176694 4.388068e-02  
## sp500 log\_pop log\_RMHI  
## nbr.val 36.000000 3.600000e+01 3.600000e+01  
## nbr.null 1.000000 0.000000e+00 0.000000e+00  
## nbr.na 0.000000 0.000000e+00 0.000000e+00  
## min -38.490000 1.396016e+01 1.073189e+01  
## max 34.110000 1.410699e+01 1.099101e+01  
## range 72.600000 1.468279e-01 2.591187e-01  
## sum 341.220000 5.061283e+02 3.920212e+02  
## median 11.895000 1.407078e+01 1.090695e+01  
## mean 9.478333 1.405912e+01 1.088948e+01  
## SE.mean 2.652485 7.430996e-03 1.018766e-02  
## CI.mean.0.95 5.384831 1.508572e-02 2.068204e-02  
## var 253.284340 1.987909e-03 3.736381e-03  
## std.dev 15.914909 4.458598e-02 6.112594e-02  
## coef.var 1.679083 3.171321e-03 5.613304e-03

The dependent variable, unemployment, has a mean of 5.27 and a standard deviation of 1.57. The lowest unemployment rate was in 2017 at 3.0, while the highest was in 2009 at 8.10. The mean HPI was 352.03, with a standard deviation of 125.66. The lowest HPI was in 1984 at 141.12, while the highest was in 2018 at 548.46. Poverty rate had a mean of 11.94 and a standard deviation of 1.41. In 1994, lowest rate was seen of 9.4, and the highest was 15.4 in 1993. The sp500 index mean was 9.48, and the standard deviation was 15.70. In 2008, the lowest index was seen at -38.4, while the highest index was seen in 1995 at 34.11. The logged population mean was 14.06, and the standard deviation was 0.045. The lowest logged population at 13.96 in 1984, and the highest in 2018 at 14.11. Logged real median household income had a mean of 10.89 and a standard deviation of 0.061. In 1985, Maine had the lowest log\_RMHI at 10.73, while had the highest in 2013 at 10.99.

#pairwise correlations  
sapply(ME\_data, class)

## year period periodName value state quarter   
## "numeric" "character" "character" "numeric" "character" "numeric"   
## annual HPI MHI RMHI poverty population   
## "numeric" "numeric" "numeric" "numeric" "numeric" "numeric"   
## sp500 log\_pop log\_RMHI   
## "numeric" "numeric" "numeric"

sapply(ME\_data, is.factor)

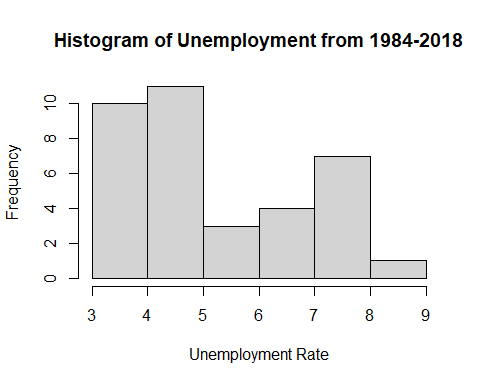
## year period periodName value state quarter annual   
## FALSE FALSE FALSE FALSE FALSE FALSE FALSE   
## HPI MHI RMHI poverty population sp500 log\_pop   
## FALSE FALSE FALSE FALSE FALSE FALSE FALSE   
## log\_RMHI   
## FALSE

cor(ME\_data[sapply(ME\_data, function(x) !is.character(x))])

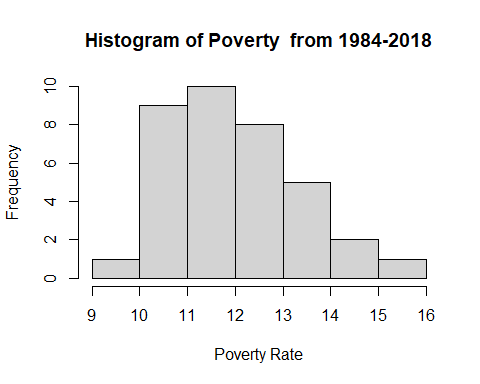
## Warning in cor(ME\_data[sapply(ME\_data, function(x) !is.character(x))]): the  
## standard deviation is zero

## year value quarter annual HPI MHI  
## year 1.00000000 -0.02315555 NA NA 0.93709957 0.98584154  
## value -0.02315555 1.00000000 NA NA -0.02225343 -0.01498802  
## quarter NA NA 1 NA NA NA  
## annual NA NA NA 1 NA NA  
## HPI 0.93709957 -0.02225343 NA NA 1.00000000 0.93912596  
## MHI 0.98584154 -0.01498802 NA NA 0.93912596 1.00000000  
## RMHI 0.62306486 -0.15868513 NA NA 0.63636855 0.72926512  
## poverty 0.09185477 0.38089057 NA NA 0.09147306 0.04393559  
## population 0.95734408 0.06660653 NA NA 0.94162902 0.94981247  
## sp500 -0.13852928 0.03233341 NA NA -0.21137251 -0.11773558  
## log\_pop 0.95348785 0.06676556 NA NA 0.93593949 0.94588350  
## log\_RMHI 0.62581314 -0.15870064 NA NA 0.63826853 0.72965027  
## RMHI poverty population sp500 log\_pop  
## year 0.62306486 0.09185477 0.95734408 -0.13852928 0.95348785  
## value -0.15868513 0.38089057 0.06660653 0.03233341 0.06676556  
## quarter NA NA NA NA NA  
## annual NA NA NA NA NA  
## HPI 0.63636855 0.09147306 0.94162902 -0.21137251 0.93593949  
## MHI 0.72926512 0.04393559 0.94981247 -0.11773558 0.94588350  
## RMHI 1.00000000 -0.29105621 0.65765201 0.04814258 0.65934566  
## poverty -0.29105621 1.00000000 0.07320758 -0.15092375 0.07023557  
## population 0.65765201 0.07320758 1.00000000 -0.17974171 0.99979500  
## sp500 0.04814258 -0.15092375 -0.17974171 1.00000000 -0.17810308  
## log\_pop 0.65934566 0.07023557 0.99979500 -0.17810308 1.00000000  
## log\_RMHI 0.99923468 -0.29175775 0.66292201 0.04104175 0.66488803  
## log\_RMHI  
## year 0.62581314  
## value -0.15870064  
## quarter NA  
## annual NA  
## HPI 0.63826853  
## MHI 0.72965027  
## RMHI 0.99923468  
## poverty -0.29175775  
## population 0.66292201  
## sp500 0.04104175  
## log\_pop 0.66488803  
## log\_RMHI 1.00000000

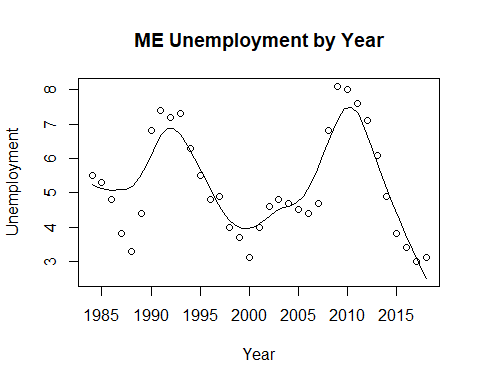
hist(ME\_data$value,  
 main="Histogram of Unemployment from 1984-2018",  
 xlab="Unemployment Rate",  
 breaks = 5)

 In Maine, from 1984-2018, the annual unemployment rate ranged from 3.0-8.9%. The histogram displays a positively skewed distribution. 18 of the 36 years were characterized by annual unemployment rates in lowest two intervals, 3.0%-3.9% & 4.0-4.9%.

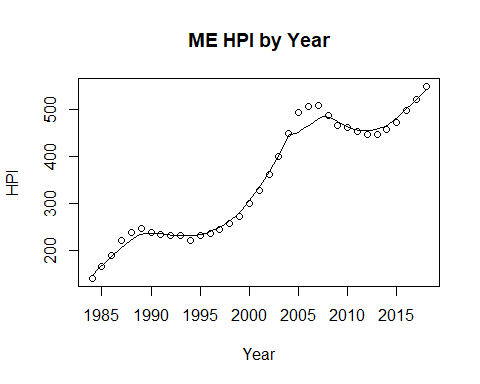
hist(ME\_data$poverty,  
 main="Histogram of Poverty from 1984-2018",  
 xlab="Poverty Rate",  
 breaks = 5)

 In Maine, from 1984-2018, the annual poverty rates ranged from 9.0%-15.9%. The distribution displayed by the histogram is not symmetrical, however, the interval at each end of the range experienced only one occurrence signifying outliers, to a degree.

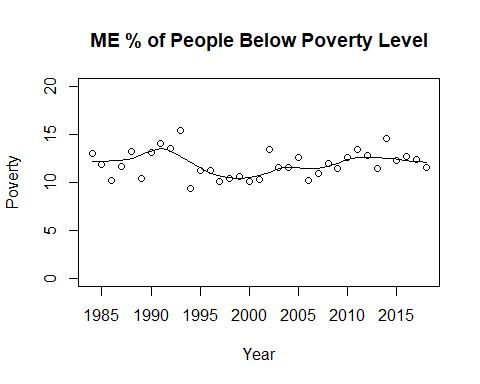
scatter.smooth(ME\_data$year, ME\_data$value, main="ME Unemployment by Year", xlab = "Year", ylab = "Unemployment", span = 2/8)



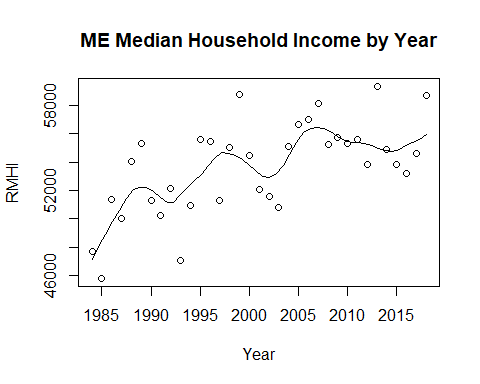
scatter.smooth(ME\_data$year, ME\_data$HPI, main="ME HPI by Year", xlab = "Year", ylab = "HPI", span = 2/8)



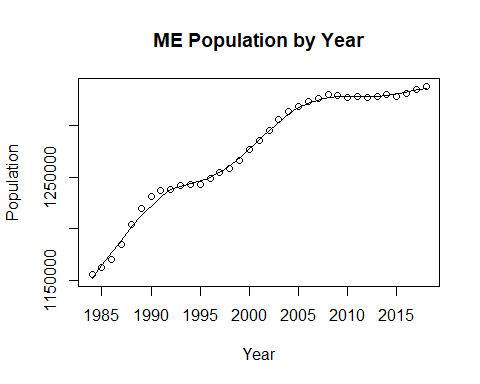
scatter.smooth(x=ME\_data$year, y=ME\_data$poverty, main="ME % of People Below Poverty Level", xlab = "Year", ylab = "Poverty", ylim = c(0,20), span = 2/8)



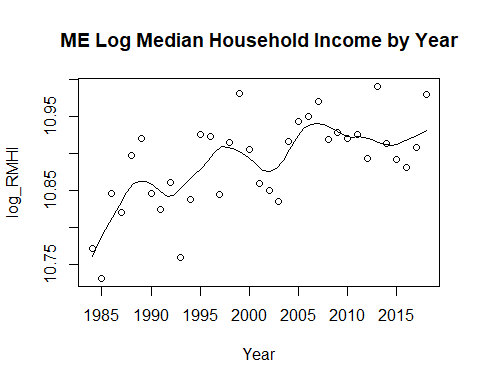
ME\_data$RMHI <- as.integer(ME\_data$RMHI)  
scatter.smooth(x=ME\_data$year, y=ME\_data$RMHI, main="ME Median Household Income by Year", xlab = "Year", ylab = "RMHI", span = 2/8 )



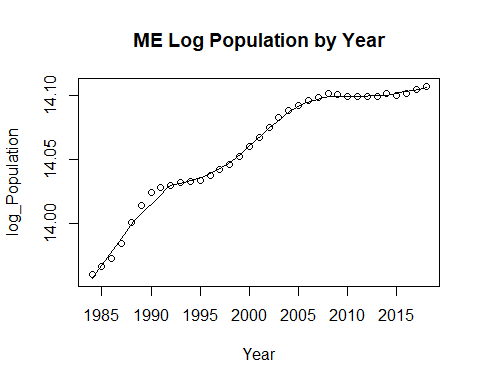
ME\_data$population <- as.integer(ME\_data$population)  
scatter.smooth(x=ME\_data$year, y=ME\_data$population, main="ME Population by Year", xlab = "Year", ylab = "Population", span = 2/8 )



#ME\_data$log\_RMHI <- as.integer(ME\_data$log\_RMHI)  
scatter.smooth(x=ME\_data$year, y=ME\_data$log\_RMHI, main="ME Log Median Household Income by Year", xlab = "Year", ylab = "log\_RMHI", span = 2/8 )



#ME\_data$log\_pop <- as.integer(ME\_data$log\_pop)  
scatter.smooth(x=ME\_data$year, y=ME\_data$log\_pop, main="ME Log Population by Year", xlab = "Year", ylab = "log\_Population", span = 2/8)



cor(ME\_data$RMHI, ME\_data$value)

## [1] -0.1586851

cor(ME\_data$HPI, ME\_data$value)

## [1] -0.02225343

cor(ME\_data$poverty, ME\_data$value)

## [1] 0.3808906

cor(ME\_data$population, ME\_data$value)

## [1] 0.06660653

cor(ME\_data$sp500, ME\_data$value)

## [1] 0.03233341

cor(ME\_data$log\_pop, ME\_data$value)

## [1] 0.06676556

cor(ME\_data$log\_RMHI, ME\_data$value)

## [1] -0.1587006

#run base correlations between categories and dependent variable

The correlation between the median household income and unemployment is Maine is -0.1586851. THe correlation between the median household income and unemployment is a small negative relationship. The negative relationship suggests that the two variables tend to move inversely of one another.

The correlation between the housing price index and unemployment in Maine is -0.02225343. The correlation between the two variables is that of a small negative relationship. The relation between the two variables is a small correlation that points to the two variables moving inversely of one another.

The correlation between poverty, the proportion of the population below the poverty line, and unemployment in Maine is 0.3808906. The correlation between poverty and unemployment is that of a medium positive correlation. The positive correlation suggests that the two variables tend to mimic one another with the direction that they proceed.

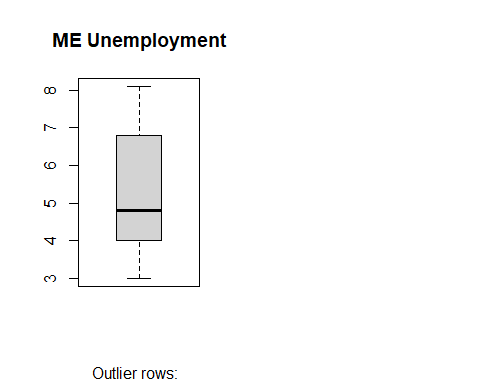
The correlation between the population and unemployment in Maine is 0.06660653. The correlation between the population and unemployment is that of a small positive correlation. A small positive correlation suggests that there is a weak positive relationship between the population and unemployment, but the two variables do tend to move in the same direction of one another.

The correlation between the S&P 500 and unemployment in Maine is 0.03233341. The correlation between the S&P 500 and unemployment is a small positive correlation. The small positive correlation is quite close to 0, but does still show the small possibility that the two variables move in the same direction of one another.

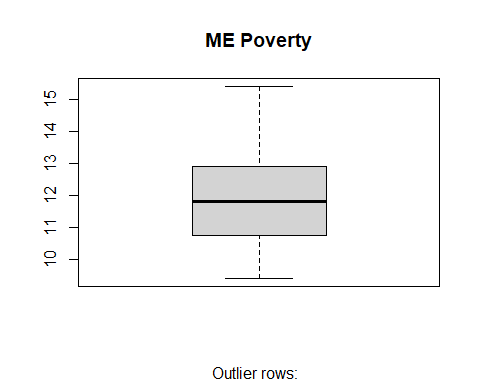
The correlation between the log of the population and unemployment in Maine is 0.06676556. The correlation between the log of the population and unemployment is quite similar to that of the S&P 500. The correlation is a small positive correlation between the two variables.

The correlation between the log of the median household income and unemployment in Maine is -0.1587006. The correlation between the log of the median household income and unemployment is that of a small negative relationship. The small negative relationship suggests that the two variables have a weak inverse relationship of one another.

par(mfrow=c(1, 2)) # divide graph area in 2 columns  
boxplot(ME\_data$value, main="ME Unemployment", sub=paste("Outlier rows: ", boxplot.stats(ME\_data$value)$out)) # box plot for 'Unemployment'



boxplot(ME\_data$poverty, main="ME Poverty", sub=paste("Outlier rows: ", boxplot.stats(ME\_data$poverty)$out)) # box plot for 'Poverty'



#create box plot for poverty level and unemployment level

#run multiple linear model for data  
ME\_reg1 <- lm(value ~ poverty + RMHI + HPI + population + sp500, data = ME\_data)  
summary(ME\_reg1)

##   
## Call:  
## lm(formula = value ~ poverty + RMHI + HPI + population + sp500,   
## data = ME\_data)  
##   
## Residuals:  
## Min 1Q Median 3Q Max   
## -2.27145 -1.16612 0.00577 1.04225 2.86079   
##   
## Coefficients:  
## Estimate Std. Error t value Pr(>|t|)   
## (Intercept) -2.373e+01 1.560e+01 -1.521 0.1386   
## poverty 4.010e-01 2.020e-01 1.985 0.0563 .  
## RMHI -7.119e-05 1.190e-04 -0.598 0.5540   
## HPI -9.596e-03 6.043e-03 -1.588 0.1228   
## population 2.453e-05 1.369e-05 1.792 0.0832 .  
## sp500 8.747e-03 1.663e-02 0.526 0.6028   
## ---  
## Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1  
##   
## Residual standard error: 1.485 on 30 degrees of freedom  
## Multiple R-squared: 0.2376, Adjusted R-squared: 0.1105   
## F-statistic: 1.869 on 5 and 30 DF, p-value: 0.1295

anova(ME\_reg1)

## Analysis of Variance Table  
##   
## Response: value  
## Df Sum Sq Mean Sq F value Pr(>F)   
## poverty 1 12.581 12.5806 5.7083 0.02337 \*  
## RMHI 1 0.217 0.2167 0.0983 0.75602   
## HPI 1 0.085 0.0855 0.0388 0.84520   
## population 1 7.107 7.1073 3.2249 0.08261 .  
## sp500 1 0.609 0.6094 0.2765 0.60285   
## Residuals 30 66.117 2.2039   
## ---  
## Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1

ME\_reg2 <- lm(value ~ HPI + poverty + sp500 + log\_pop + log\_RMHI, data = ME\_data)  
summary(ME\_reg2)

##   
## Call:  
## lm(formula = value ~ HPI + poverty + sp500 + log\_pop + log\_RMHI,   
## data = ME\_data)  
##   
## Residuals:  
## Min 1Q Median 3Q Max   
## -2.28170 -1.16703 -0.01773 1.04684 2.87688   
##   
## Coefficients:  
## Estimate Std. Error t value Pr(>|t|)   
## (Intercept) -3.604e+02 2.243e+02 -1.607 0.1186   
## HPI -8.837e-03 5.791e-03 -1.526 0.1375   
## poverty 3.996e-01 2.029e-01 1.969 0.0582 .  
## sp500 8.832e-03 1.665e-02 0.531 0.5996   
## log\_pop 2.893e+01 1.660e+01 1.743 0.0916 .  
## log\_RMHI -3.938e+00 6.334e+00 -0.622 0.5388   
## ---  
## Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1  
##   
## Residual standard error: 1.489 on 30 degrees of freedom  
## Multiple R-squared: 0.2334, Adjusted R-squared: 0.1057   
## F-statistic: 1.827 on 5 and 30 DF, p-value: 0.1376

anova(ME\_reg2)

## Analysis of Variance Table  
##   
## Response: value  
## Df Sum Sq Mean Sq F value Pr(>F)   
## HPI 1 0.043 0.0429 0.0194 0.89021   
## poverty 1 12.823 12.8227 5.7871 0.02252 \*  
## sp500 1 0.569 0.5690 0.2568 0.61605   
## log\_pop 1 5.952 5.9524 2.6864 0.11165   
## log\_RMHI 1 0.857 0.8566 0.3866 0.53879   
## Residuals 30 66.473 2.2158   
## ---  
## Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1

ME\_linearModelSignificant <- lm(value ~ poverty + log\_pop, data = ME\_data)  
summary(ME\_linearModelSignificant)

##   
## Call:  
## lm(formula = value ~ poverty + log\_pop, data = ME\_data)  
##   
## Residuals:  
## Min 1Q Median 3Q Max   
## -2.5280 -1.1369 -0.1391 1.1325 3.0014   
##   
## Coefficients:  
## Estimate Std. Error t value Pr(>|t|)   
## (Intercept) -19.7465 79.8885 -0.247 0.8063   
## poverty 0.4235 0.1806 2.345 0.0252 \*  
## log\_pop 1.4196 5.6910 0.249 0.8046   
## ---  
## Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1  
##   
## Residual standard error: 1.497 on 33 degrees of freedom  
## Multiple R-squared: 0.1467, Adjusted R-squared: 0.09497   
## F-statistic: 2.836 on 2 and 33 DF, p-value: 0.07299

Next, the data we collected is considered a time series as it was measured on an annual basis. We used the unemployment data that was collected for each state and implemented a time series forecast using the Holt Method.

Time series data typically contains four components: trend, seasonal, cyclical, and random components. Trend is represented by the long-term movements in the data series, whether that be upward or downward. A seasonal component normally represents repetitions that occur within a single year time period, whereas the cyclical component represents long-term trends that are usually a factor the economy. Typically, seasonal trends are easier to identify than cyclical trends because the time period of the seasonal trend is often known ahead of time. Random components represent the random and unexplained movements within the time series data.

As our unemployment data is shown on an annual basis, we will be focusing on the trend, cyclical, random components in our analysis. The type of analysis we have chosen to focus on is labeled the Holt Exponential Smoothing Method. This method is characterized by incorporating the large upward and downward fluctuations in the series and is most appropriate when the dataset lacks seasonal variation. We went into greater detail explaining the process that was used on IL, but for the other states we will simply analyze the model and the results.

IL\_TS <- ts(IL\_data$value, start = c(1984), end = c(2018), frequency = 1)  
#Stores start, end, and frequency of timeseries data  
  
IL\_TData <- window(IL\_TS, end = c(2004))  
IL\_VData <- window(IL\_TS, start = c(2005))  
#Partitions the training and validation sets

In the above section of code, we begin our time series analysis of IL by creating the time series with the TS function. We set the start and end values of our times series data equal to the time span our data streches and set the frequency equal to one.

Using the window function we partition our dataset into a training set, IL\_TData, and a validation set, IL\_VData.

IL\_HUser <- ets(IL\_TData, model = "AAN", alpha = 0.2, beta = 0.15)  
summary(IL\_HUser)

## ETS(A,Ad,N)   
##   
## Call:  
## ets(y = IL\_TData, model = "AAN", alpha = 0.2, beta = 0.15)   
##   
## Smoothing parameters:  
## alpha = 0.2   
## beta = 0.15   
## phi = 0.8   
##   
## Initial states:  
## l = 9.1038   
## b = -0.492   
##   
## sigma: 1.3058  
##   
## AIC AICc BIC   
## 77.43215 79.93215 81.61024   
##   
## Training set error measures:  
## ME RMSE MAE MPE MAPE MASE ACF1  
## Training set 0.01280128 1.139837 0.8591663 -1.525969 14.65093 1.26348 0.5841472

#Training data set using user based parameters for alpha and beta for smoothing.

It is common for practitioners to provide user supplied values for alpha and beta (smoothing parameters) due to the fact that the computer generated values are known to over-fit models where the data performed well in the sample period, but does not exhibit the same performance in the future. In our data, we will compare both user supplied and computer generated smoothing parameters. We will then compare the error measures to determine which model is a better fit to our data. We used an alpha = 0.2 and beta = 0.15 when creating our user supplier parameters model.

The ets function above denotes the error, trend, and seasonality of the training set, respectively. The first letter in the model string represents that we want the error to be additive, the second letter represents that fact that we expect the trend type to be additive, and the third letter represents the fact that we do not intend on incorporating seasonality into our model.

IL\_HCmp <- ets(IL\_TData, model = "AAN")  
summary(IL\_HCmp)

## ETS(A,A,N)   
##   
## Call:  
## ets(y = IL\_TData, model = "AAN")   
##   
## Smoothing parameters:  
## alpha = 0.9999   
## beta = 1e-04   
##   
## Initial states:  
## l = 8.3553   
## b = -0.2338   
##   
## sigma: 0.9784  
##   
## AIC AICc BIC   
## 68.58121 72.58121 73.80382   
##   
## Training set error measures:  
## ME RMSE MAE MPE MAPE MASE  
## Training set 0.1120952 0.8803193 0.6387882 1.287488 10.53478 0.9393944  
## ACF1  
## Training set 0.0652965

#Training data set using computer based parameters for alpha and beta for smoothing.

By comparing the root mean squared error(RMSE) and the mean absolute percentage error(MAPE) we can assess the fitness of the user supplier and computer generate models. The user supplied model has a higher RMSE and MAPE indicating that the computer generated model provides a better overall fit to our forecast.

IL\_nV <- length(IL\_VData)  
IL\_fUser <- forecast(IL\_HUser, h = IL\_nV)  
IL\_fCmp <- forecast(IL\_HCmp, h = IL\_nV)

The length function above was used to set the number of observations equal to the length of the validation set. Then the forecast function was used to make a forecast based on the observations in our training set.

accuracy(IL\_fUser,IL\_VData)

## ME RMSE MAE MPE MAPE MASE  
## Training set 0.01280128 1.139837 0.8591663 -1.525969 14.65093 1.263480  
## Test set 0.22556242 2.075490 1.8905876 -5.598974 27.98416 2.780276  
## ACF1 Theil's U  
## Training set 0.5841472 NA  
## Test set 0.6727078 1.402801

accuracy(IL\_fCmp,IL\_VData)

## ME RMSE MAE MPE MAPE MASE  
## Training set 0.1120952 0.8803193 0.6387882 1.287488 10.53478 0.9393944  
## Test set 3.0944149 3.8292855 3.3229011 37.967956 43.12811 4.8866193  
## ACF1 Theil's U  
## Training set 0.0652965 NA  
## Test set 0.6934401 2.220317

When comparing the accuracy function results we see conflicting results in our validation set. The test set shows that the RMSE and MAPE of the user generated model provides a better fit for the performance of our data. Ultimately, we decided to continue the process using the computer generated model, but the user supplied model would work just as well.

IL\_HFinal <- ets(IL\_TS, model = "AAN")  
forecast(IL\_HFinal, h=1)

## Point Forecast Lo 80 Hi 80 Lo 95 Hi 95  
## 2019 4.41789 2.851058 5.984722 2.021628 6.814152

Finally, we bring everything together with the ets function and forecast the unemployment level for the next period, 2019. The forecast that was generated shows us that the expected value of the Illinois unemployment level in December 2019 is equal to 4.42% based on the trends that were observed in our model.

TX\_TS <- ts(TX\_data$value, start = c(1984), end = c(2018), frequency = 1)  
#Stores start, end, and frequency of timeseries data  
  
TX\_TData <- window(TX\_TS, end = c(2004))  
TX\_VData <- window(TX\_TS, start = c(2005))  
#Partitions the training and validation sets

TX\_HUser <- ets(TX\_TData, model = "AAN", alpha = 0.2, beta = 0.15)  
summary(TX\_HUser)

## ETS(A,Ad,N)   
##   
## Call:  
## ets(y = TX\_TData, model = "AAN", alpha = 0.2, beta = 0.15)   
##   
## Smoothing parameters:  
## alpha = 0.2   
## beta = 0.15   
## phi = 0.8   
##   
## Initial states:  
## l = 7.0129   
## b = 0.1951   
##   
## sigma: 1.1717  
##   
## AIC AICc BIC   
## 72.87881 75.37881 77.05690   
##   
## Training set error measures:  
## ME RMSE MAE MPE MAPE MASE  
## Training set -0.06394251 1.022727 0.8466247 -2.613638 14.79038 1.200886  
## ACF1  
## Training set 0.476818

#Training data set using user based parameters for alpha and beta for smoothing.

TX\_HCmp <- ets(TX\_TData, model = "AAN")  
summary(TX\_HCmp)

## ETS(A,A,N)   
##   
## Call:  
## ets(y = TX\_TData, model = "AAN")   
##   
## Smoothing parameters:  
## alpha = 0.9999   
## beta = 1e-04   
##   
## Initial states:  
## l = 5.7864   
## b = 0.0087   
##   
## sigma: 0.9299  
##   
## AIC AICc BIC   
## 66.44290 70.44290 71.66551   
##   
## Training set error measures:  
## ME RMSE MAE MPE MAPE MASE  
## Training set -0.02238698 0.8366222 0.6733549 -1.470508 11.35325 0.9551133  
## ACF1  
## Training set 0.1427012

#Training data set using computer based parameters for alpha and beta for smoothing.

The user supplied model for the state of Texas has a higher RMSE and MAPE at 1.0227 and 14.7904 respectively. The computer generated model has a RMSE of 0.8366 and a MAPE of 11.3532 indicating that the computer generated model provides a better overall fit to our forecast.

TX\_nV <- length(TX\_VData)  
TX\_fUser <- forecast(TX\_HUser, h = TX\_nV)  
TX\_fCmp <- forecast(TX\_HCmp, h = TX\_nV)

accuracy(TX\_fUser,TX\_VData)

## ME RMSE MAE MPE MAPE MASE  
## Training set -0.06394251 1.022727 0.8466247 -2.613638 14.79038 1.200886  
## Test set -0.78836930 1.689115 1.5022550 -21.986608 31.15916 2.130858  
## ACF1 Theil's U  
## Training set 0.4768180 NA  
## Test set 0.7280731 1.949414

accuracy(TX\_fCmp,TX\_VData)

## ME RMSE MAE MPE MAPE MASE  
## Training set -0.02238698 0.8366222 0.6733549 -1.470508 11.35325 0.9551133  
## Test set -0.02241831 1.4831870 1.2931856 -7.302041 24.09310 1.8343058  
## ACF1 Theil's U  
## Training set 0.1427012 NA  
## Test set 0.7360214 1.480138

The accuracy function confirms that the computer generated model is a better overall fit to the data we are using to create our model. This is due to the fact that once again the validation set exhibits a lower RMSE and MAPE.

TX\_HFinal <- ets(TX\_TS, model = "AAN")  
forecast(TX\_HFinal, h=1)

## Point Forecast Lo 80 Hi 80 Lo 95 Hi 95  
## 2019 3.788149 2.577206 4.999093 1.936171 5.640127

The forecast that was generated shows us that the expected value of the Texas unemployment level in December 2019 is equal to 3.79% based on the trends that were observed in our model.

MO\_TS <- ts(MO\_data$value, start = c(1984), end = c(2018), frequency = 1)  
#Stores start, end, and frequency of timeseries data  
  
MO\_TData <- window(MO\_TS, end = c(2004))  
MO\_VData <- window(MO\_TS, start = c(2005))  
#Partitions the training and validation sets

MO\_HUser <- ets(MO\_TData, model = "AAN", alpha = 0.2, beta = 0.15)  
summary(MO\_HUser)

## ETS(A,Ad,N)   
##   
## Call:  
## ets(y = MO\_TData, model = "AAN", alpha = 0.2, beta = 0.15)   
##   
## Smoothing parameters:  
## alpha = 0.2   
## beta = 0.15   
## phi = 0.8   
##   
## Initial states:  
## l = 5.7779   
## b = -0.0413   
##   
## sigma: 1.0464  
##   
## AIC AICc BIC   
## 68.13046 70.63046 72.30855   
##   
## Training set error measures:  
## ME RMSE MAE MPE MAPE MASE  
## Training set 0.06931804 0.9133979 0.7747159 -0.6806828 16.60718 1.324301  
## ACF1  
## Training set 0.6113182

#Training data set using user based parameters for alpha and beta for smoothing.

MO\_HCmp <- ets(MO\_TData, model = "AAN")  
summary(MO\_HCmp)

## ETS(A,A,N)   
##   
## Call:  
## ets(y = MO\_TData, model = "AAN")   
##   
## Smoothing parameters:  
## alpha = 0.9998   
## beta = 1e-04   
##   
## Initial states:  
## l = 6.4123   
## b = -0.0902   
##   
## sigma: 0.7288  
##   
## AIC AICc BIC   
## 56.20810 60.20810 61.43071   
##   
## Training set error measures:  
## ME RMSE MAE MPE MAPE MASE  
## Training set 0.05636837 0.6556892 0.5531963 0.2900369 11.59243 0.9456347  
## ACF1  
## Training set 0.01565671

#Training data set using computer based parameters for alpha and beta for smoothing.

The user supplied model for the state of Missouri has a higher RMSE and MAPE at 0.9134 and 16.6072 respectively. The computer generated model has a RMSE of 0.6557 and a MAPE of 11.5924 indicating that the computer generated model provides a better overall fit to our forecast.

MO\_nV <- length(MO\_VData)  
MO\_fUser <- forecast(MO\_HUser, h = MO\_nV)  
MO\_fCmp <- forecast(MO\_HCmp, h = MO\_nV)

accuracy(MO\_fUser,MO\_VData)

## ME RMSE MAE MPE MAPE MASE  
## Training set 0.06931804 0.9133979 0.7747159 -0.6806828 16.60718 1.324301  
## Test set -0.09513439 2.0611563 1.7152490 -13.2943741 31.72038 2.932050  
## ACF1 Theil's U  
## Training set 0.6113182 NA  
## Test set 0.7831950 2.287588

accuracy(MO\_fCmp,MO\_VData)

## ME RMSE MAE MPE MAPE MASE  
## Training set 0.05636837 0.6556892 0.5531963 0.2900369 11.59243 0.9456347  
## Test set 1.13313797 2.2176982 1.7394279 10.1296118 25.23538 2.9733810  
## ACF1 Theil's U  
## Training set 0.01565671 NA  
## Test set 0.75380126 1.730281

The results from the accuracy function conflict with our initial assessment. The accuracy function shows that in the validation set, the RMSE and MAPE of the user defined model are lower and indicate there is a better overall fit to the data. For this exercise we will proceed with the computer generated model.

MO\_HFinal <- ets(MO\_TS, model = "AAN")  
forecast(MO\_HFinal, h=1)

## Point Forecast Lo 80 Hi 80 Lo 95 Hi 95  
## 2019 2.942054 1.799567 4.084541 1.194771 4.689337

The forecast that was generated shows us that the expected value of the Missouri unemployment level in December 2019 is equal to 2.94% based on the trends that were observed in our model.

WA\_TS <- ts(WA\_data$value, start = c(1984), end = c(2018), frequency = 1)  
#Stores start, end, and frequency of timeseries data  
  
WA\_TData <- window(WA\_TS, end = c(2004))  
WA\_VData <- window(WA\_TS, start = c(2005))  
#Partitions the training and validation sets

WA\_HUser <- ets(WA\_TData, model = "AAN", alpha = 0.2, beta = 0.15)  
summary(WA\_HUser)

## ETS(A,Ad,N)   
##   
## Call:  
## ets(y = WA\_TData, model = "AAN", alpha = 0.2, beta = 0.15)   
##   
## Smoothing parameters:  
## alpha = 0.2   
## beta = 0.15   
## phi = 0.8   
##   
## Initial states:  
## l = 9.4939   
## b = -1.1554   
##   
## sigma: 1.327  
##   
## AIC AICc BIC   
## 78.10616 80.60616 82.28425   
##   
## Training set error measures:  
## ME RMSE MAE MPE MAPE MASE ACF1  
## Training set 0.1668724 1.158276 0.9222942 0.3641993 14.66641 1.289922 0.5982543

#Training data set using user based parameters for alpha and beta for smoothing.

WA\_HCmp <- ets(WA\_TData, model = "AAN")  
summary(WA\_HCmp)

## ETS(A,A,N)   
##   
## Call:  
## ets(y = WA\_TData, model = "AAN")   
##   
## Smoothing parameters:  
## alpha = 0.9899   
## beta = 1e-04   
##   
## Initial states:  
## l = 9.1822   
## b = -0.1524   
##   
## sigma: 0.9284  
##   
## AIC AICc BIC   
## 66.37627 70.37627 71.59889   
##   
## Training set error measures:  
## ME RMSE MAE MPE MAPE MASE  
## Training set -0.00335368 0.8352961 0.6750692 -0.5337995 10.6124 0.9441527  
## ACF1  
## Training set 0.2760247

#Training data set using computer based parameters for alpha and beta for smoothing.

The user supplied model for the state of Washington has a higher RMSE and MAPE at 1.1583 and 14.6664 respectively. The computer generated model has a RMSE of 0.8353 and a MAPE of 10.6124 indicating that the computer generated model provides a better overall fit to our forecast.

WA\_nV <- length(WA\_VData)  
WA\_fUser <- forecast(WA\_HUser, h = WA\_nV)  
WA\_fCmp <- forecast(WA\_HCmp, h = WA\_nV)

accuracy(WA\_fUser,WA\_VData)

## ME RMSE MAE MPE MAPE MASE  
## Training set 0.1668724 1.158276 0.9222942 0.3641993 14.66641 1.289922  
## Test set -0.5459622 2.013298 1.8266143 -16.3512750 29.49131 2.554705  
## ACF1 Theil's U  
## Training set 0.5982543 NA  
## Test set 0.7158472 1.60662

accuracy(WA\_fCmp,WA\_VData)

## ME RMSE MAE MPE MAPE MASE  
## Training set -0.00335368 0.8352961 0.6750692 -0.5337995 10.61240 0.9441527  
## Test set 2.05500222 2.8605907 2.3427478 24.7608910 30.59774 3.2765703  
## ACF1 Theil's U  
## Training set 0.2760247 NA  
## Test set 0.7089211 1.777545

The results from the accuracy function conflict with our initial assessment. The accuracy function shows that in the validation set, the RMSE and MAPE of the user defined model are lower and indicate there is a better overall fit to the data. For this exercise we will proceed with the computer generated model.

WA\_HFinal <- ets(WA\_TS, model = "AAN")  
forecast(WA\_HFinal, h=1)

## Point Forecast Lo 80 Hi 80 Lo 95 Hi 95  
## 2019 4.62788 3.183544 6.072216 2.418959 6.836801

The forecast that was generated shows us that the expected value of the Washington unemployment level in December 2019 is equal to 4.63% based on the trends that were observed in our model.

OR\_TS <- ts(OR\_data$value, start = c(1984), end = c(2018), frequency = 1)  
#Stores start, end, and frequency of timeseries data  
  
OR\_TData <- window(OR\_TS, end = c(2004))  
OR\_VData <- window(OR\_TS, start = c(2005))  
#Partitions the training and validation sets

OR\_HUser <- ets(OR\_TData, model = "AAN", alpha = 0.2, beta = 0.15)  
summary(OR\_HUser)

## ETS(A,Ad,N)   
##   
## Call:  
## ets(y = OR\_TData, model = "AAN", alpha = 0.2, beta = 0.15)   
##   
## Smoothing parameters:  
## alpha = 0.2   
## beta = 0.15   
## phi = 0.8   
##   
## Initial states:  
## l = 9.9304   
## b = -1.7166   
##   
## sigma: 1.388  
##   
## AIC AICc BIC   
## 79.99563 82.49563 84.17372   
##   
## Training set error measures:  
## ME RMSE MAE MPE MAPE MASE ACF1  
## Training set 0.3044074 1.211574 0.9415815 2.433037 14.9369 1.101265 0.4619036

#Training data set using user based parameters for alpha and beta for smoothing.

OR\_HCmp <- ets(OR\_TData, model = "AAN")  
summary(OR\_HCmp)

## ETS(A,A,N)   
##   
## Call:  
## ets(y = OR\_TData, model = "AAN")   
##   
## Smoothing parameters:  
## alpha = 0.9999   
## beta = 0.0366   
##   
## Initial states:  
## l = 9.7897   
## b = -0.2329   
##   
## sigma: 1.1935  
##   
## AIC AICc BIC   
## 76.92740 80.92740 82.15002   
##   
## Training set error measures:  
## ME RMSE MAE MPE MAPE MASE  
## Training set 0.09499493 1.073848 0.8136491 0.918828 13.29897 0.9516364  
## ACF1  
## Training set 0.1586164

#Training data set using computer based parameters for alpha and beta for smoothing.

The user supplied model for the state of Oregon has a higher RMSE and MAPE at 1.2116 and 14.9369 respectively. The computer generated model has a RMSE of 1.0738 and a MAPE of 13.2990 indicating that the computer generated model provides a better overall fit to our forecast.

OR\_nV <- length(OR\_VData)  
OR\_fUser <- forecast(OR\_HUser, h = OR\_nV)  
OR\_fCmp <- forecast(OR\_HCmp, h = OR\_nV)

accuracy(OR\_fUser,OR\_VData)

## ME RMSE MAE MPE MAPE MASE  
## Training set 0.3044074 1.211574 0.9415815 2.433037 14.93690 1.101265  
## Test set -0.7593753 2.487879 2.2496261 -24.662687 39.73627 2.631142  
## ACF1 Theil's U  
## Training set 0.4619036 NA  
## Test set 0.7605760 1.984834

accuracy(OR\_fCmp,OR\_VData)

## ME RMSE MAE MPE MAPE MASE  
## Training set 0.09499493 1.073848 0.8136491 0.918828 13.29897 0.9516364  
## Test set 1.85661552 2.915292 2.3169545 18.443408 27.83266 2.7098883  
## ACF1 Theil's U  
## Training set 0.1586164 NA  
## Test set 0.7384358 1.479877

The results from the accuracy function conflict with our initial assessment. The accuracy function shows that in the validation set, the RMSE of the user defined model is lower and indicate there is a better overall fit to the data. The MAPE of the validation set confirms our initial assessment that the computer generated model provides a better overall fit. For this exercise we will proceed with the computer generated model.

OR\_HFinal <- ets(OR\_TS, model = "AAN")  
forecast(OR\_HFinal, h=1)

## Point Forecast Lo 80 Hi 80 Lo 95 Hi 95  
## 2019 3.767832 2.080512 5.455152 1.187299 6.348366

The forecast that was generated shows us that the expected value of the Oregon unemployment level in December 2019 is equal to 3.7678% based on the trends that were observed in our model.

CO\_TS <- ts(CO\_data$value, start = c(1984), end = c(2018), frequency = 1)  
#Stores start, end, and frequency of timeseries data  
  
CO\_TData <- window(CO\_TS, end = c(2004))  
CO\_VData <- window(CO\_TS, start = c(2005))  
#Partitions the training and validation sets

CO\_HUser <- ets(CO\_TData, model = "AAN", alpha = 0.2, beta = 0.15)  
summary(CO\_HUser)

## ETS(A,A,N)   
##   
## Call:  
## ets(y = CO\_TData, model = "AAN", alpha = 0.2, beta = 0.15)   
##   
## Smoothing parameters:  
## alpha = 0.2   
## beta = 0.15   
##   
## Initial states:  
## l = 7.304   
## b = 0.102   
##   
## sigma: 1.3863  
##   
## AIC AICc BIC   
## 79.21659 80.62836 82.35016   
##   
## Training set error measures:  
## ME RMSE MAE MPE MAPE MASE ACF1  
## Training set 0.2453149 1.247312 0.9271383 4.558743 18.29914 1.211945 0.4497532

#Training data set using user based parameters for alpha and beta for smoothing.

CO\_HCmp <- ets(CO\_TData, model = "AAN")  
summary(CO\_HCmp)

## ETS(A,A,N)   
##   
## Call:  
## ets(y = CO\_TData, model = "AAN")   
##   
## Smoothing parameters:  
## alpha = 0.9477   
## beta = 1e-04   
##   
## Initial states:  
## l = 5.7459   
## b = 0.0425   
##   
## sigma: 1.0929  
##   
## AIC AICc BIC   
## 73.22703 77.22703 78.44964   
##   
## Training set error measures:  
## ME RMSE MAE MPE MAPE MASE  
## Training set -0.07075571 0.9832854 0.7630149 -3.657183 15.53447 0.9974051  
## ACF1  
## Training set 0.1307753

#Training data set using computer based parameters for alpha and beta for smoothing.

The user supplied model for the state of Colorado has a higher RMSE and MAPE at 1.2473 and 18.2991 respectively. The computer generated model has a RMSE of 0.9833 and a MAPE of 15.5345 indicating that the computer generated model provides a better overall fit to our forecast.

CO\_nV <- length(CO\_VData)  
CO\_fUser <- forecast(CO\_HUser, h = CO\_nV)  
CO\_fCmp <- forecast(CO\_HCmp, h = CO\_nV)

accuracy(CO\_fUser,CO\_VData)

## ME RMSE MAE MPE MAPE MASE  
## Training set 0.2453149 1.247312 0.9271383 4.558743 18.29914 1.211945  
## Test set -6.2702267 7.788098 6.2702267 -163.078186 163.07819 8.196375  
## ACF1 Theil's U  
## Training set 0.4497532 NA  
## Test set 0.7902830 10.08042

accuracy(CO\_fCmp,CO\_VData)

## ME RMSE MAE MPE MAPE MASE  
## Training set -0.07075571 0.9832854 0.7630149 -3.657183 15.53447 0.9974051  
## Test set 0.01828570 2.2206448 1.9569048 -17.858301 42.02405 2.5580454  
## ACF1 Theil's U  
## Training set 0.1307753 NA  
## Test set 0.8034183 2.32387

The accuracy function confirms that the computer generated model is a better overall fit to the data we are using to create our model. This is due to the fact that once again the validation set exhibits a lower RMSE and MAPE.

CO\_HFinal <- ets(CO\_TS, model = "AAN")  
forecast(CO\_HFinal, h=1)

## Point Forecast Lo 80 Hi 80 Lo 95 Hi 95  
## 2019 2.724606 1.31318 4.136033 0.5660158 4.883197

The forecast that was generated shows us that the expected value of the Colorado unemployment level in December 2019 is equal to 2.7246% based on the trends that were observed in our model.

WV\_TS <- ts(WV\_data$value, start = c(1984), end = c(2018), frequency = 1)  
#Stores start, end, and frequency of timeseries data  
  
WV\_TData <- window(WV\_TS, end = c(2004))  
WV\_VData <- window(WV\_TS, start = c(2005))  
#Partitions the training and validation sets

WV\_HUser <- ets(WV\_TData, model = "AAN", alpha = 0.2, beta = 0.15)  
summary(WV\_HUser)

## ETS(A,Ad,N)   
##   
## Call:  
## ets(y = WV\_TData, model = "AAN", alpha = 0.2, beta = 0.15)   
##   
## Smoothing parameters:  
## alpha = 0.2   
## beta = 0.15   
## phi = 0.8   
##   
## Initial states:  
## l = 12.7771   
## b = -1.5609   
##   
## sigma: 1.6369  
##   
## AIC AICc BIC   
## 86.92211 89.42211 91.10020   
##   
## Training set error measures:  
## ME RMSE MAE MPE MAPE MASE ACF1  
## Training set -0.1320942 1.428802 1.148726 -3.883647 14.15979 1.228584 0.504785

#Training data set using user based parameters for alpha and beta for smoothing.

WV\_HCmp <- ets(WV\_TData, model = "AAN")  
summary(WV\_HCmp)

## ETS(A,Ad,N)   
##   
## Call:  
## ets(y = WV\_TData, model = "AAN")   
##   
## Smoothing parameters:  
## alpha = 0.9999   
## beta = 2e-04   
## phi = 0.8417   
##   
## Initial states:  
## l = 15.9638   
## b = -1.8651   
##   
## sigma: 1.1154  
##   
## AIC AICc BIC   
## 74.81012 80.81012 81.07726   
##   
## Training set error measures:  
## ME RMSE MAE MPE MAPE MASE  
## Training set -0.07675973 0.9735733 0.6947987 -1.828428 8.590638 0.7431002  
## ACF1  
## Training set 0.271826

#Training data set using computer based parameters for alpha and beta for smoothing.

The user supplied model for the state of West Virginia has a higher RMSE and MAPE at 1.4288 and 14.1598 respectively. The computer generated model has a RMSE of 0.9736 and a MAPE of 8.5906 indicating that the computer generated model provides a better overall fit to our forecast.

WV\_nV <- length(WV\_VData)  
WV\_fUser <- forecast(WV\_HUser, h = WV\_nV)  
WV\_fCmp <- forecast(WV\_HCmp, h = WV\_nV)

accuracy(WV\_fUser,WV\_VData)

## ME RMSE MAE MPE MAPE MASE ACF1  
## Training set -0.1320942 1.428802 1.148726 -3.883647 14.15979 1.228584 0.5047850  
## Test set 1.7441115 2.355068 1.857933 23.491691 26.16469 1.987094 0.7058261  
## Theil's U  
## Training set NA  
## Test set 1.672258

accuracy(WV\_fCmp,WV\_VData)

## ME RMSE MAE MPE MAPE MASE  
## Training set -0.07675973 0.9735733 0.6947987 -1.828428 8.590638 0.7431002  
## Test set 1.62448769 2.2608082 1.7776928 21.469084 25.065397 1.9012758  
## ACF1 Theil's U  
## Training set 0.2718260 NA  
## Test set 0.7051023 1.604304

The accuracy function confirms that the computer generated model is a better overall fit to the data we are using to create our model. This is due to the fact that once again the validation set exhibits a lower RMSE and MAPE.

WV\_HFinal <- ets(WV\_TS, model = "AAN")  
forecast(WV\_HFinal, h=1)

## Point Forecast Lo 80 Hi 80 Lo 95 Hi 95  
## 2019 4.699914 3.200919 6.198909 2.407398 6.99243

The forecast that was generated shows us that the expected value of the West Virginia unemployment level in December 2019 is equal to 4.70% based on the trends that were observed in our model.

ME\_TS <- ts(ME\_data$value, start = c(1984), end = c(2018), frequency = 1)  
#Stores start, end, and frequency of timeseries data  
  
ME\_TData <- window(ME\_TS, end = c(2004))  
ME\_VData <- window(ME\_TS, start = c(2005))  
#Partitions the training and validation sets

ME\_HUser <- ets(ME\_TData, model = "AAN", alpha = 0.2, beta = 0.15)  
summary(ME\_HUser)

## ETS(A,Ad,N)   
##   
## Call:  
## ets(y = ME\_TData, model = "AAN", alpha = 0.2, beta = 0.15)   
##   
## Smoothing parameters:  
## alpha = 0.2   
## beta = 0.15   
## phi = 0.8   
##   
## Initial states:  
## l = 4.0432   
## b = 0.2976   
##   
## sigma: 1.6532  
##   
## AIC AICc BIC   
## 87.33868 89.83868 91.51677   
##   
## Training set error measures:  
## ME RMSE MAE MPE MAPE MASE  
## Training set -0.02033974 1.443044 1.294447 -4.593562 26.58756 2.022574  
## ACF1  
## Training set 0.7185094

#Training data set using user based parameters for alpha and beta for smoothing.

ME\_HCmp <- ets(ME\_TData, model = "AAN")  
summary(ME\_HCmp)

## ETS(A,A,N)   
##   
## Call:  
## ets(y = ME\_TData, model = "AAN")   
##   
## Smoothing parameters:  
## alpha = 0.9999   
## beta = 0.0927   
##   
## Initial states:  
## l = 5.2667   
## b = -0.4686   
##   
## sigma: 0.9636  
##   
## AIC AICc BIC   
## 67.94131 71.94131 73.16392   
##   
## Training set error measures:  
## ME RMSE MAE MPE MAPE MASE  
## Training set 0.2000609 0.8670087 0.6127466 3.267401 12.13418 0.9574165  
## ACF1  
## Training set 0.4424009

#Training data set using computer based parameters for alpha and beta for smoothing.

The user supplied model for the state of Maine has a higher RMSE and MAPE at 1.4430 and 26.5876 respectively. The computer generated model has a RMSE of 0.8670 and a MAPE of 12.1342 indicating that the computer generated model provides a better overall fit to our forecast.

ME\_nV <- length(ME\_VData)  
ME\_fUser <- forecast(ME\_HUser, h = ME\_nV)  
ME\_fCmp <- forecast(ME\_HCmp, h = ME\_nV)

accuracy(ME\_fUser,ME\_VData)

## ME RMSE MAE MPE MAPE MASE  
## Training set -0.02033974 1.443044 1.294447 -4.593562 26.58756 2.022574  
## Test set 1.12028865 2.111426 1.735987 10.222169 28.91356 2.712480  
## ACF1 Theil's U  
## Training set 0.7185094 NA  
## Test set 0.7819364 2.047897

accuracy(ME\_fCmp,ME\_VData)

## ME RMSE MAE MPE MAPE MASE  
## Training set 0.2000609 0.8670087 0.6127466 3.267401 12.13418 0.9574165  
## Test set 1.6351809 2.3593767 1.7967102 21.302558 26.12643 2.8073597  
## ACF1 Theil's U  
## Training set 0.4424009 NA  
## Test set 0.7717380 2.015639

The accuracy function confirms that the computer generated model is a better overall fit to the data we are using to create our model. This is due to the fact that once again the validation set exhibits a lower RMSE and MAPE.

ME\_HFinal <- ets(ME\_TS, model = "AAN")  
forecast(ME\_HFinal, h=1)

## Point Forecast Lo 80 Hi 80 Lo 95 Hi 95  
## 2019 2.679319 1.659277 3.69936 1.1193 4.239338

The forecast that was generated shows us that the expected value of the Maine unemployment level in December 2019 is equal to 2.6793% based on the trends that were observed in our model.