climate data - San Francisco

Allie Hsu

12/3/2019

## Temperature Changes in San Francisco

San Francisco temperature changes since 1921 cannot be attributed to becoming a heat island, as it has been developed for over a century

#file.choose()  
#read.csv("/home/CAMPUS/achb2017/Climate\_Change\_Narratives/student\_folders/Hsu/AllieHsu\_SanFrancisco\_Data.csv")  
climate\_data <- read.csv("/home/CAMPUS/achb2017/Climate\_Change\_Narratives/student\_folders/Hsu/AllieHsu\_SanFrancisco\_Data.csv")  
  
head(climate\_data)

## STATION NAME DATE DAPR MDPR PRCP PSUN  
## 1 USW00023272 SAN FRANCISCO DOWNTOWN, CA US 7/1/1921 NA NA 0 NA  
## 2 USW00023272 SAN FRANCISCO DOWNTOWN, CA US 7/2/1921 NA NA 0 NA  
## 3 USW00023272 SAN FRANCISCO DOWNTOWN, CA US 7/3/1921 NA NA 0 NA  
## 4 USW00023272 SAN FRANCISCO DOWNTOWN, CA US 7/4/1921 NA NA 0 NA  
## 5 USW00023272 SAN FRANCISCO DOWNTOWN, CA US 7/5/1921 NA NA 0 NA  
## 6 USW00023272 SAN FRANCISCO DOWNTOWN, CA US 7/6/1921 NA NA 0 NA  
## SNOW SNWD TAVG TMAX TMIN TOBS TSUN WDFM WSF1 WSFG WSFM WT01 WT03 WT04  
## 1 0 NA NA 76 54 NA NA NA NA NA NA NA NA NA  
## 2 0 NA NA 85 60 NA NA NA NA NA NA NA NA NA  
## 3 0 NA NA 85 62 NA NA NA NA NA NA NA NA NA  
## 4 0 NA NA 64 53 NA NA NA NA NA NA NA NA NA  
## 5 0 NA NA 82 52 NA NA NA NA NA NA NA NA NA  
## 6 0 NA NA 94 61 NA NA NA NA NA NA NA NA NA  
## WT05 WT08 WT11 WT16 WT18  
## 1 NA NA NA NA NA  
## 2 NA NA NA NA NA  
## 3 NA NA NA NA NA  
## 4 NA NA NA NA NA  
## 5 NA NA NA NA NA  
## 6 NA NA NA NA NA

str(climate\_data)

## 'data.frame': 35930 obs. of 26 variables:  
## $ STATION: Factor w/ 1 level "USW00023272": 1 1 1 1 1 1 1 1 1 1 ...  
## $ NAME : Factor w/ 1 level "SAN FRANCISCO DOWNTOWN, CA US": 1 1 1 1 1 1 1 1 1 1 ...  
## $ DATE : Factor w/ 35930 levels "1/1/1922","1/1/1923",..: 26823 27912 29001 29298 29397 29496 29595 29694 29793 26922 ...  
## $ DAPR : int NA NA NA NA NA NA NA NA NA NA ...  
## $ MDPR : num NA NA NA NA NA NA NA NA NA NA ...  
## $ PRCP : num 0 0 0 0 0 0 0 0 0 0 ...  
## $ PSUN : int NA NA NA NA NA NA NA NA NA NA ...  
## $ SNOW : num 0 0 0 0 0 0 0 0 0 0 ...  
## $ SNWD : int NA NA NA NA NA NA NA NA NA NA ...  
## $ TAVG : int NA NA NA NA NA NA NA NA NA NA ...  
## $ TMAX : int 76 85 85 64 82 94 90 79 64 60 ...  
## $ TMIN : int 54 60 62 53 52 61 61 56 54 52 ...  
## $ TOBS : int NA NA NA NA NA NA NA NA NA NA ...  
## $ TSUN : int NA NA NA NA NA NA NA NA NA NA ...  
## $ WDFM : int NA NA NA NA NA NA NA NA NA NA ...  
## $ WSF1 : num NA NA NA NA NA NA NA NA NA NA ...  
## $ WSFG : num NA NA NA NA NA NA NA NA NA NA ...  
## $ WSFM : num NA NA NA NA NA NA NA NA NA NA ...  
## $ WT01 : int NA NA NA NA NA NA NA NA NA NA ...  
## $ WT03 : int NA NA NA NA NA NA NA NA NA NA ...  
## $ WT04 : int NA NA NA NA NA NA NA NA NA NA ...  
## $ WT05 : int NA NA NA NA NA NA NA NA NA NA ...  
## $ WT08 : int NA NA NA NA NA NA NA NA NA NA ...  
## $ WT11 : int NA NA NA NA NA NA NA NA NA NA ...  
## $ WT16 : int NA NA NA NA NA NA NA NA NA NA ...  
## $ WT18 : int NA NA NA NA NA NA NA NA NA NA ...

names(climate\_data)

## [1] "STATION" "NAME" "DATE" "DAPR" "MDPR" "PRCP" "PSUN"   
## [8] "SNOW" "SNWD" "TAVG" "TMAX" "TMIN" "TOBS" "TSUN"   
## [15] "WDFM" "WSF1" "WSFG" "WSFM" "WT01" "WT03" "WT04"   
## [22] "WT05" "WT08" "WT11" "WT16" "WT18"

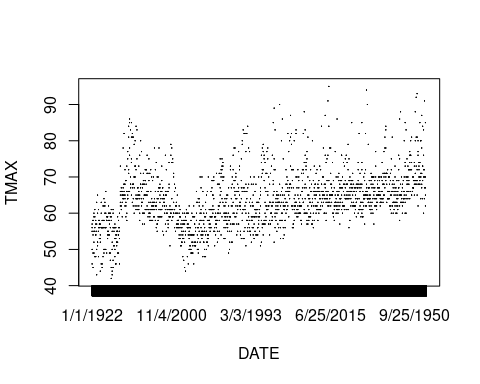
## plot(TMAX~DATE, climate\_data)  
  
min(climate\_data$TMAX)

## [1] NA

min(climate\_data$TMAX, na.rm = T)

## [1] 35

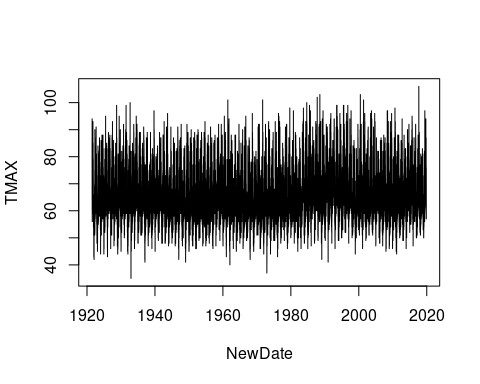
climate\_data$TMAX[climate\_data$TMAX==-9999] = NA  
climate\_data$TMIN[climate\_data$TMIN==-9999] = NA ## eliminating placeholders which will skew our averages  
  
plot(TMAX~DATE, climate\_data[1:1835,], ty='l')



str(climate\_data)

## 'data.frame': 35930 obs. of 26 variables:  
## $ STATION: Factor w/ 1 level "USW00023272": 1 1 1 1 1 1 1 1 1 1 ...  
## $ NAME : Factor w/ 1 level "SAN FRANCISCO DOWNTOWN, CA US": 1 1 1 1 1 1 1 1 1 1 ...  
## $ DATE : Factor w/ 35930 levels "1/1/1922","1/1/1923",..: 26823 27912 29001 29298 29397 29496 29595 29694 29793 26922 ...  
## $ DAPR : int NA NA NA NA NA NA NA NA NA NA ...  
## $ MDPR : num NA NA NA NA NA NA NA NA NA NA ...  
## $ PRCP : num 0 0 0 0 0 0 0 0 0 0 ...  
## $ PSUN : int NA NA NA NA NA NA NA NA NA NA ...  
## $ SNOW : num 0 0 0 0 0 0 0 0 0 0 ...  
## $ SNWD : int NA NA NA NA NA NA NA NA NA NA ...  
## $ TAVG : int NA NA NA NA NA NA NA NA NA NA ...  
## $ TMAX : int 76 85 85 64 82 94 90 79 64 60 ...  
## $ TMIN : int 54 60 62 53 52 61 61 56 54 52 ...  
## $ TOBS : int NA NA NA NA NA NA NA NA NA NA ...  
## $ TSUN : int NA NA NA NA NA NA NA NA NA NA ...  
## $ WDFM : int NA NA NA NA NA NA NA NA NA NA ...  
## $ WSF1 : num NA NA NA NA NA NA NA NA NA NA ...  
## $ WSFG : num NA NA NA NA NA NA NA NA NA NA ...  
## $ WSFM : num NA NA NA NA NA NA NA NA NA NA ...  
## $ WT01 : int NA NA NA NA NA NA NA NA NA NA ...  
## $ WT03 : int NA NA NA NA NA NA NA NA NA NA ...  
## $ WT04 : int NA NA NA NA NA NA NA NA NA NA ...  
## $ WT05 : int NA NA NA NA NA NA NA NA NA NA ...  
## $ WT08 : int NA NA NA NA NA NA NA NA NA NA ...  
## $ WT11 : int NA NA NA NA NA NA NA NA NA NA ...  
## $ WT16 : int NA NA NA NA NA NA NA NA NA NA ...  
## $ WT18 : int NA NA NA NA NA NA NA NA NA NA ...

strDates <- as.character(climate\_data$DATE)  
climate\_data$NewDate <- as.Date(strDates, "%m/%d/%Y") ##reformatting dates - NewDate  
  
plot(TMAX~NewDate, climate\_data, ty='l') ## plot of daily TMAX data



## Analyse Climate Data

null hypothesis = slope of best fit line is zero, y-intercept is zero

lm(TMAX ~ NewDate, data=climate\_data)

##   
## Call:  
## lm(formula = TMAX ~ NewDate, data = climate\_data)  
##   
## Coefficients:  
## (Intercept) NewDate   
## 6.372e+01 5.835e-05

summary(lm(TMAX ~ NewDate, data=climate\_data))

##   
## Call:  
## lm(formula = TMAX ~ NewDate, data = climate\_data)  
##   
## Residuals:  
## Min 1Q Median 3Q Max   
## -27.929 -4.935 -0.867 3.795 41.265   
##   
## Coefficients:  
## Estimate Std. Error t value Pr(>|t|)   
## (Intercept) 6.372e+01 4.010e-02 1588.8 <2e-16 \*\*\*  
## NewDate 5.835e-05 3.865e-06 15.1 <2e-16 \*\*\*  
## ---  
## Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1  
##   
## Residual standard error: 7.6 on 35926 degrees of freedom  
## (2 observations deleted due to missingness)  
## Multiple R-squared: 0.006303, Adjusted R-squared: 0.006275   
## F-statistic: 227.9 on 1 and 35926 DF, p-value: < 2.2e-16

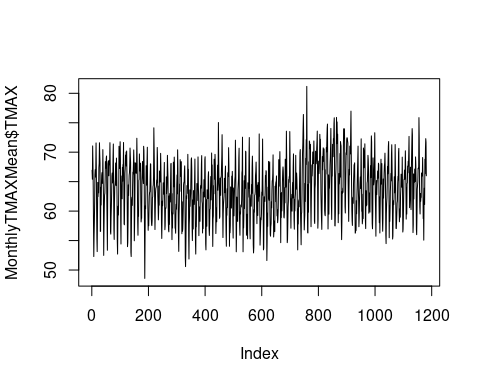
## slightly positive temperature slope (5.835e-05) - highly significant (2e-16 - is much less than 0.05)  
## lots of variability, not date driven

we reject the null hypothesis!

## disaggregate into month and year variables  
climate\_data$Month = format(as.Date(climate\_data$NewDate), format = "%m")  
climate\_data$Year = format(climate\_data$NewDate, format="%Y")  
  
MonthlyTMAXMean = aggregate(TMAX ~ Month + Year, climate\_data, mean)  
MonthlyTMAXMean$YEAR = as.numeric(MonthlyTMAXMean$Year)  
MonthlyTMAXMean$MONTH = as.numeric(MonthlyTMAXMean$Month)  
  
str(MonthlyTMAXMean)

## 'data.frame': 1181 obs. of 5 variables:  
## $ Month: chr "07" "08" "09" "10" ...  
## $ Year : chr "1921" "1921" "1921" "1921" ...  
## $ TMAX : num 66.9 65.4 71.1 68.1 64.1 ...  
## $ YEAR : num 1921 1921 1921 1921 1921 ...  
## $ MONTH: num 7 8 9 10 11 12 1 2 3 4 ...

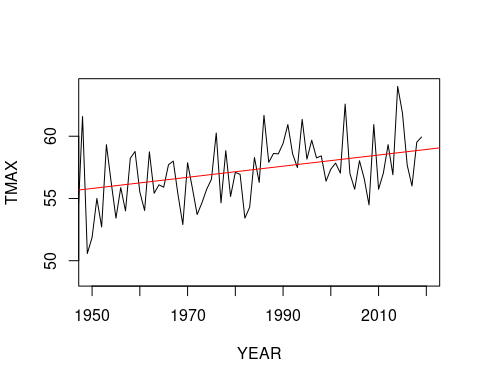
plot(MonthlyTMAXMean$TMAX, ty='l')



plot(TMAX~YEAR, data=MonthlyTMAXMean[MonthlyTMAXMean$Month=="01",],  
ty='l', xlim=c(1950, 2020))  
January.lm <- lm(TMAX~YEAR, data=MonthlyTMAXMean[MonthlyTMAXMean$Month=="01",])  
summary(January.lm)

##   
## Call:  
## lm(formula = TMAX ~ YEAR, data = MonthlyTMAXMean[MonthlyTMAXMean$Month ==   
## "01", ])  
##   
## Residuals:  
## Min 1Q Median 3Q Max   
## -6.6428 -1.5843 -0.2245 1.5296 5.8649   
##   
## Coefficients:  
## Estimate Std. Error t value Pr(>|t|)   
## (Intercept) -31.473738 16.507514 -1.907 0.0596 .   
## YEAR 0.044758 0.008376 5.343 6.14e-07 \*\*\*  
## ---  
## Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1  
##   
## Residual standard error: 2.346 on 96 degrees of freedom  
## Multiple R-squared: 0.2292, Adjusted R-squared: 0.2212   
## F-statistic: 28.55 on 1 and 96 DF, p-value: 6.137e-07

abline(coef(January.lm), col="red")



MonthlyTMINMean = aggregate(TMIN ~ Month + Year, climate\_data, mean)  
MonthlyTMINMean$YEAR = as.numeric(MonthlyTMINMean$Year)  
  
MonthlyTMINMean$YEAR = as.numeric(MonthlyTMINMean$Year)  
MonthlyTMINMean$MONTH = as.numeric(MonthlyTMINMean$Month)  
head(MonthlyTMINMean)

## Month Year TMIN YEAR MONTH  
## 1 07 1921 52.64516 1921 7  
## 2 08 1921 53.67742 1921 8  
## 3 09 1921 55.50000 1921 9  
## 4 10 1921 54.87097 1921 10  
## 5 11 1921 51.50000 1921 11  
## 6 12 1921 48.09677 1921 12

plot(MonthlyTMINMean$TMIN, ty='l')

