

## Global Land Temperatures for San Francisco

### Load the preprocessed libraries as needed.

```
library(ggplot2)

## Warning: package 'ggplot2' was built under R version 3.3.2

library(zoo)

## Warning: package 'zoo' was built under R version 3.3.2

##
## Attaching package: 'zoo'

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

library(randomForest)

## randomForest 4.6-12

## Type rfNews() to see new features/changes/bug fixes.

##
## Attaching package: 'randomForest'

## The following object is masked from 'package:ggplot2':
##
##   margin

library(data.table)
```

### Let's load the temperature data by the city, Let's say San Francisco

```
GlobalLandTemperaturesByCity <-
read.csv('./GlobalLandTemperaturesByCity.csv')
GlobalLandTemperaturesByCity <-
fread("./GlobalLandTemperaturesByCity.csv")

##
Read 6.2% of 8599212 rows
Read 12.7% of 8599212 rows
Read 19.2% of 8599212 rows
Read 26.4% of 8599212 rows
Read 32.9% of 8599212 rows
Read 40.1% of 8599212 rows
Read 47.8% of 8599212 rows
```

```
Read 54.5% of 8599212 rows
Read 61.4% of 8599212 rows
Read 68.1% of 8599212 rows
Read 74.9% of 8599212 rows
Read 82.0% of 8599212 rows
Read 88.3% of 8599212 rows
Read 95.0% of 8599212 rows
Read 8599212 rows and 7 (of 7) columns from 0.496 GB file in 00:00:16
```

## Let's Try to Use the San Francisco Data to get some insights.

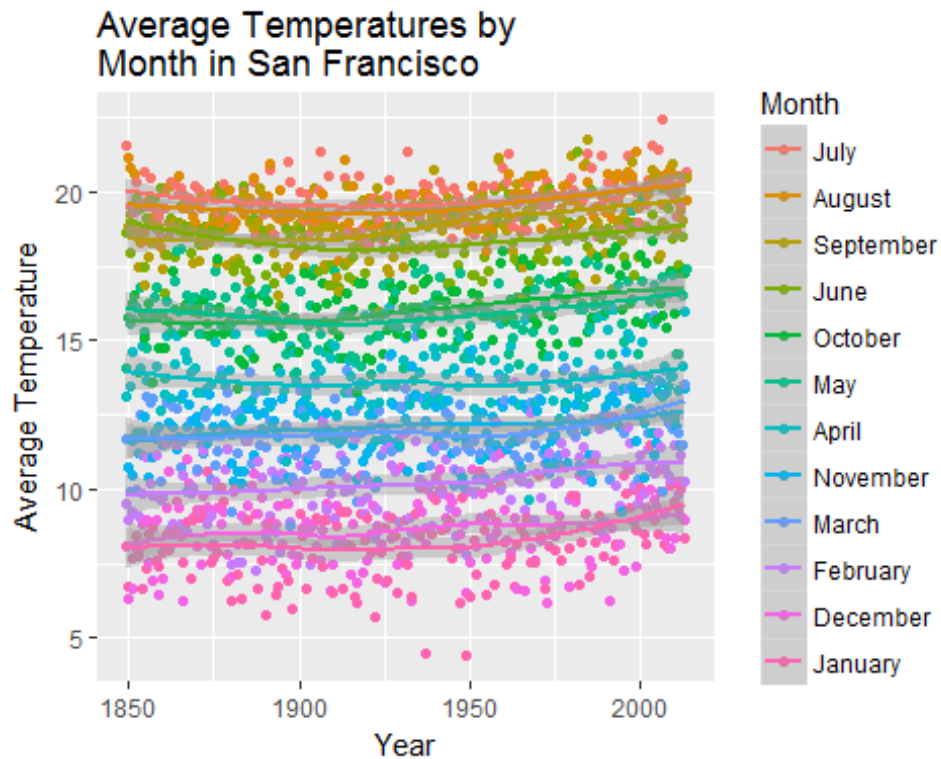
```
san.francisco<-na.omit(subset(GlobalLandTemperaturesByCity, City=="San
Francisco"))
san.francisco$dt<-as.Date(san.francisco$dt, "%Y-%m-%d")
san.francisco$lat<-as.numeric(gsub("N|E|S|W",
"", san.francisco$Latitude))*ifelse(grepl("S", san.francisco$Latitude), -
1, 1)
san.francisco$long<-as.numeric(gsub("N|E|S|W",
"", san.francisco$Longitude))*ifelse(grepl("W", san.francisco$Longitude),
-1, 1)
san.francisco$Month<-as.numeric(format(san.francisco$dt, "%m"))
san.francisco$Month.String<-format(san.francisco$dt, "%B")
san.francisco$Year<-as.numeric(format(san.francisco$dt, "%Y"))
san.francisco$elevation<-
with(san.francisco, sunPosition(as.numeric(format(dt, "%Y")), as.numeric(f
ormat(dt, "%m")), 1, 12, 0, 0, lat, long)$elevation)
san.francisco$azimuth<-
with(san.francisco, sunPosition(as.numeric(format(dt, "%Y")), as.numeric(f
ormat(dt, "%m")), 1, 12, 0, 0, lat, long)$azimuth)
```

## Graphing the Temperatures:

The graph below demonstrates the temperature categorizing from the hottest to the coldest in months. It is based on the last 150 plus years.

```
ggplot(san.francisco, aes(x=dt, y=AverageTemperature, color=reorder(Month.
String, -AverageTemperature, mean)))+
  geom_point()+geom_smooth()+ggtitle("Average Temperatures by\nMonth in
San Francisco")+
  xlab("Year")+ylab("Average Temperature")+labs(color='Month')

## `geom_smooth()` using method = 'loess'
```



```
rm(mean)
```

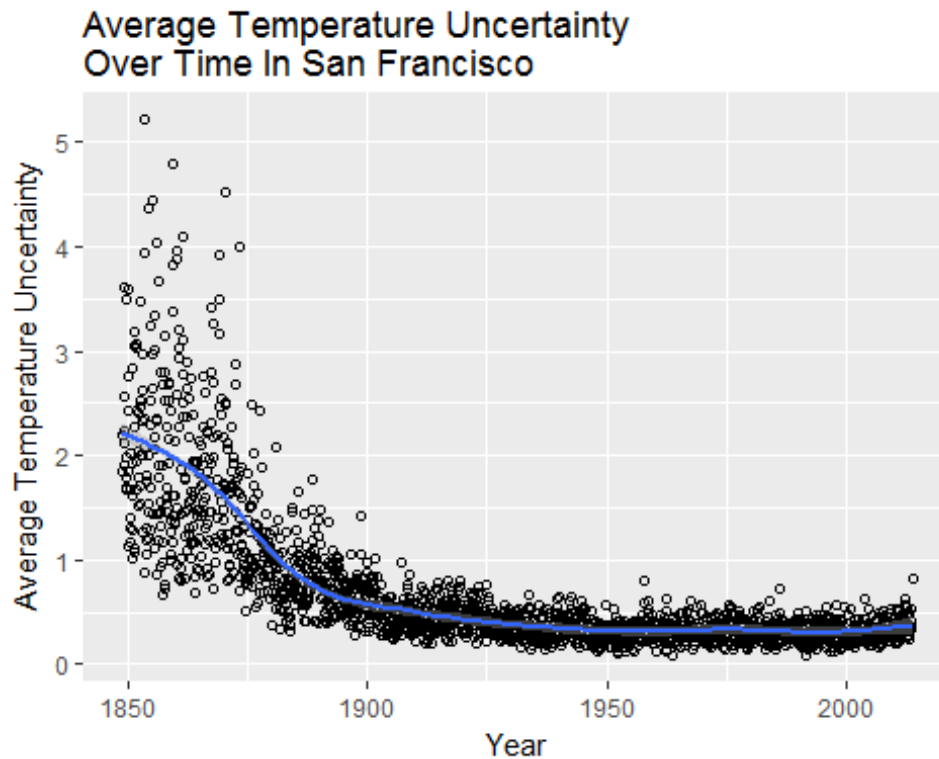
```
## Warning in rm(mean): object 'mean' not found
```

## Temperaure Uncertainty

The graph shows the declining temperature for the last 150 years plus.

```
ggplot(san.francisco,aes(x=dt,y=AverageTemperatureUncertainty))+
  geom_point(shape=1)+geom_smooth()+ggtitle("Average Temperature
Uncertainty\nOver Time In San Francisco")+
  xlab("Year")+ylab("Average Temperature Uncertainty")
```

```
## `geom_smooth()` using method = 'gam'
```

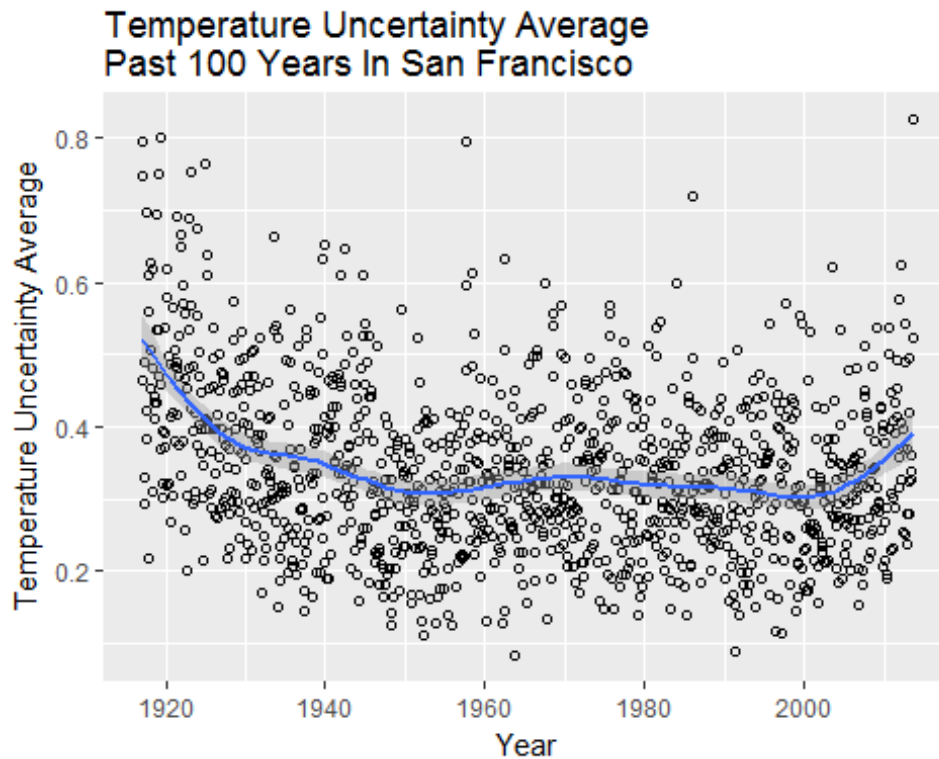


## Temperature Uncertainty 100 Years Ago

Demonstrating the accuracy for the last 100 years ago.

```
ggplot(san.francisco[Year>1916,],aes(x=dt,y=AverageTemperatureUncertainty))+
  geom_point(shape=1)+geom_smooth()+ggtitle("Temperature Uncertainty
Average\nPast 100 Years In San Francisco")+
  xlab("Year")+ylab("Temperature Uncertainty Average")

## `geom_smooth()` using method = 'gam'
```

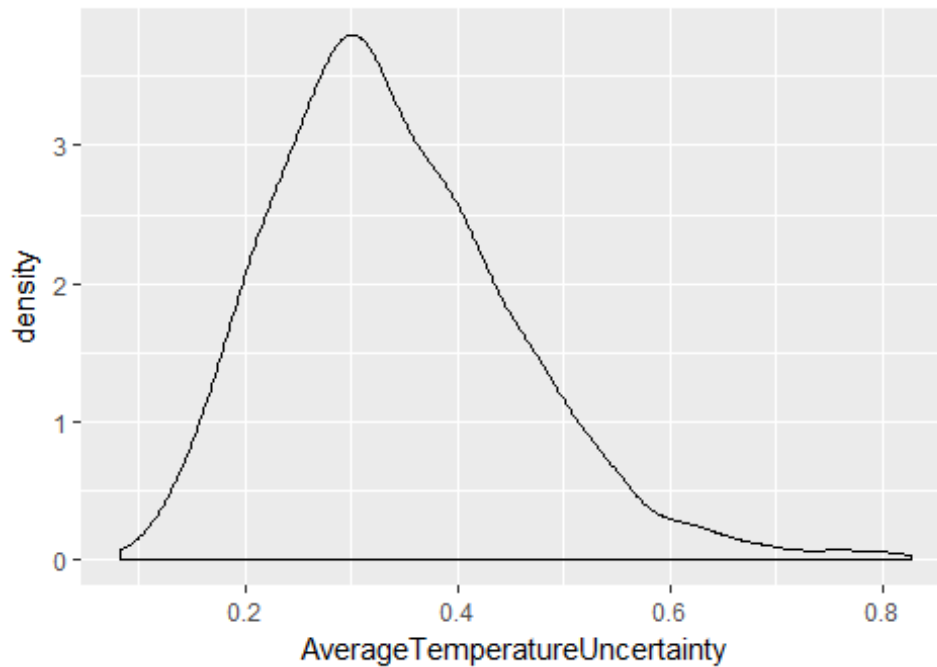


## Density Plot

In this section, the graph demonstrates the distribution on the temperature uncertainty for the last 100 years in SF.

```
ggplot(san.francisco[Year>1916,], aes(x=AverageTemperatureUncertainty))
+ geom_density()+
  ggtitle("Density Plot of Temperature Uncertainty Average\npast 100
years in San Francisco")
```

Density Plot of Temperature Uncertainty Average  
past 100 years in San Francisco



## Random Forest

```
rf<-  
randomForest(subset(san.francisco,select=c(Year,elevation,azimuth,AverageTemperatureUncertainty)),san.francisco$AverageTemperature)  
varImpPlot(rf,main="Variable Importance in Determining\nSan Francisco  
Average Temperatures")
```

## Variable Importance in Determining San Francisco Average Temperature

