Global Land Temperatures for Oakland

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 Oakland

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

```
##
Read 6.7% of 8599212 rows
Read 13.3% of 8599212 rows
Read 19.9% of 8599212 rows
Read 26.5% of 8599212 rows
Read 33.6% of 8599212 rows
Read 40.6% of 8599212 rows
Read 47.2% of 8599212 rows
```

```
Read 54.1% of 8599212 rows
Read 60.8% of 8599212 rows
Read 67.8% of 8599212 rows
Read 74.7% of 8599212 rows
Read 80.9% of 8599212 rows
Read 88.1% 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
```

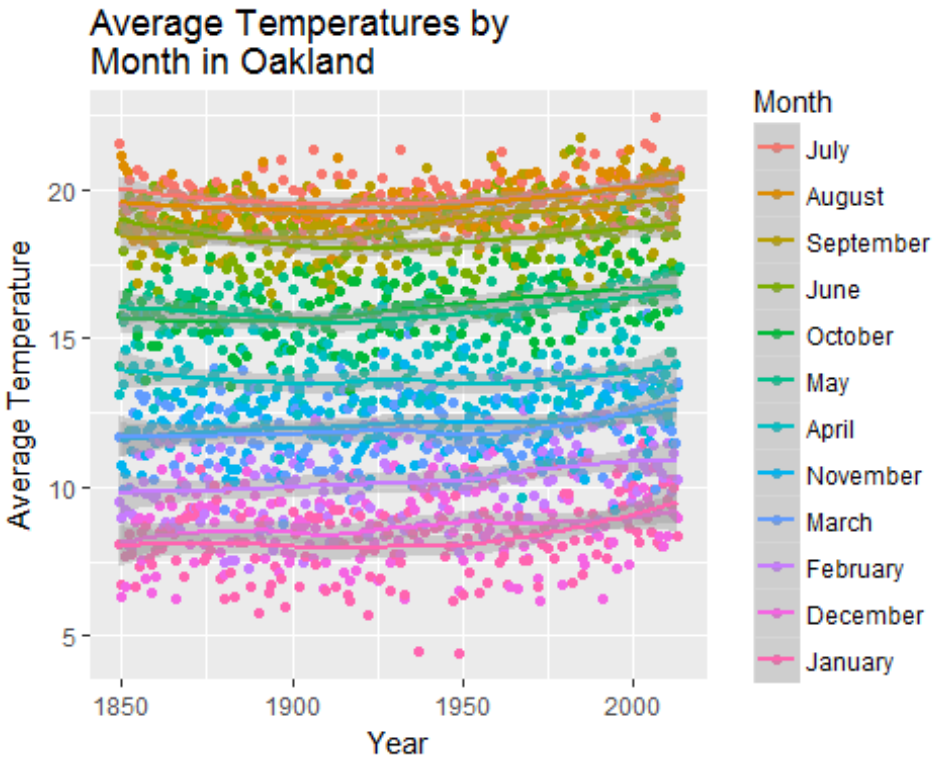
Let's Try to Use the Oakland Data to get some insights.

```
oakland<-na.omit(subset(GlobalLandTemperaturesByCity,City=="Oakland"))
oakland$dt<-as.Date(oakland$dt,"%Y-%m-%d")
oakland$lat<-as.numeric(gsub("N|E|S|W",
"",oakland$Latitude))*ifelse(grepl("S",oakland$Latitude),-1,1)
oakland$long<-as.numeric(gsub("N|E|S|W",
"",oakland$Longitude))*ifelse(grepl("W",oakland$Longitude),-1,1)
oakland$Month<-as.numeric(format(oakland$dt,"%m"))
oakland$Month.String<-format(oakland$dt,"%B")
oakland$Year<-as.numeric(format(oakland$dt,"%Y"))
oakland$elevation<-
with(oakland,sunPosition(as.numeric(format(dt,"%Y")),as.numeric(format(
dt,"%m")),1,12,0,0,lat,long)$elevation)
oakland$azimuth<-
with(oakland,sunPosition(as.numeric(format(dt,"%Y")),as.numeric(format(
dt,"%m")),1,12,0,0,lat,long)$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(oakland,aes(x=dt,y=AverageTemperature,color=reorder(Month.String
,-AverageTemperature,mean)))+
  geom_point()+geom_smooth()+ggtitle("Average Temperatures by\nMonth in
Oakland")+
  xlab("Year")+ylab("Average Temperature")+labs(color='Month')

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

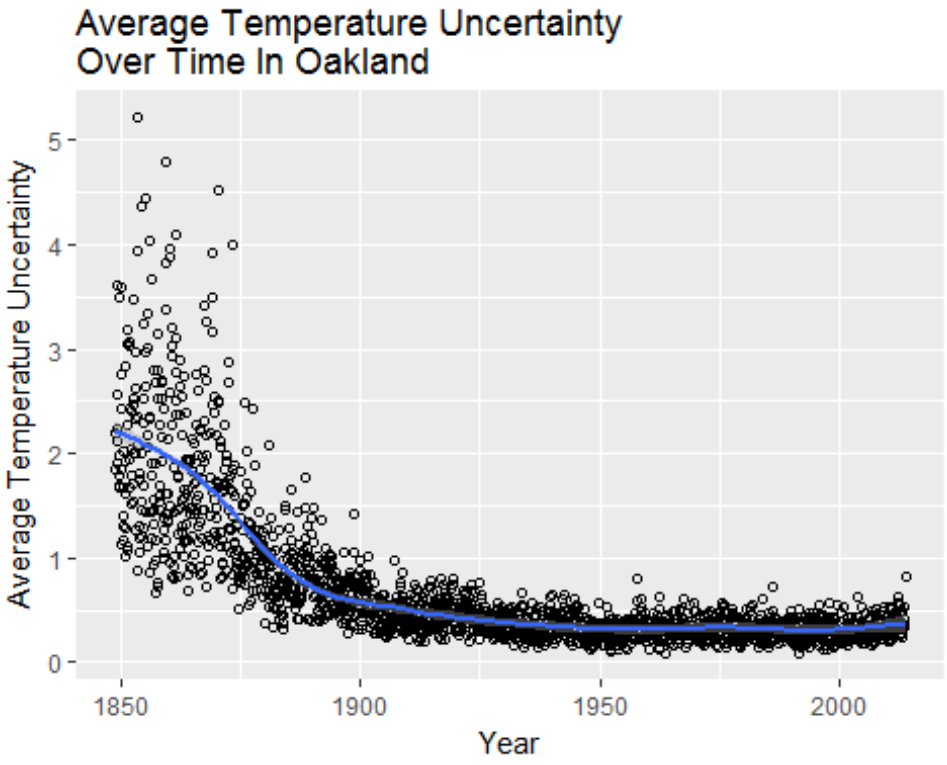



```
rm(mean)

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

Temperaure Uncertainty

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

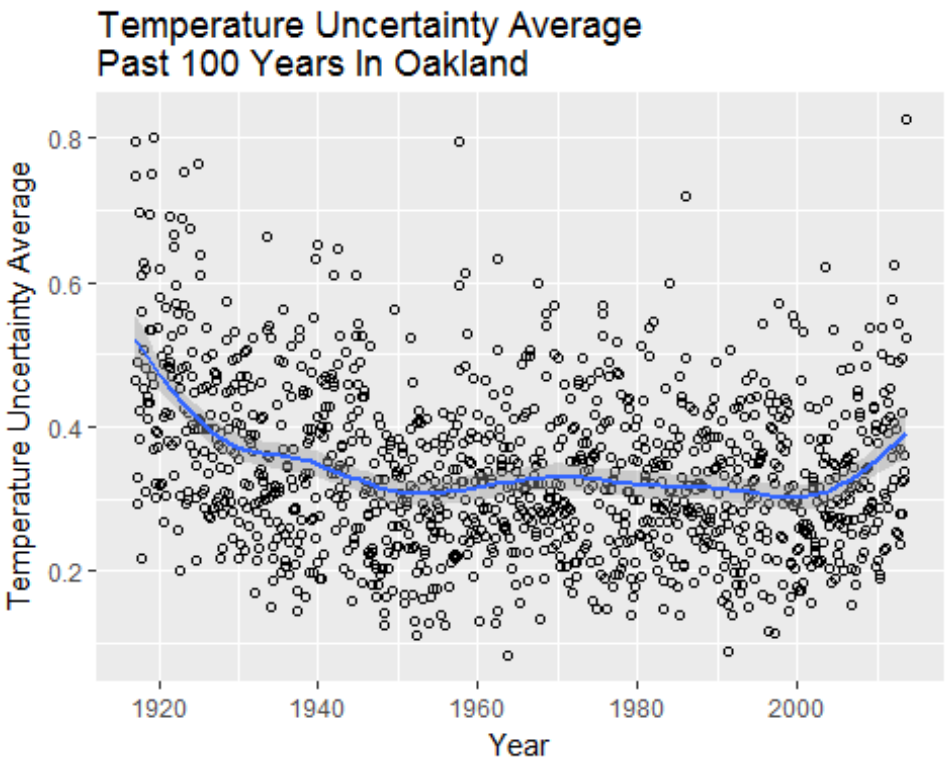
```
ggplot(oakland,aes(x=dt,y=AverageTemperatureUncertainty))+
  geom_point(shape=1)+geom_smooth()+ggtitle("Average Temperature
Uncertainty\nOver Time In Oakland")+
  xlab("Year")+ylab("Average Temperature Uncertainty")

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



Temperature Uncertainty 100 Years Ago

Demonstrating the accuracy for the last 100 years ago.

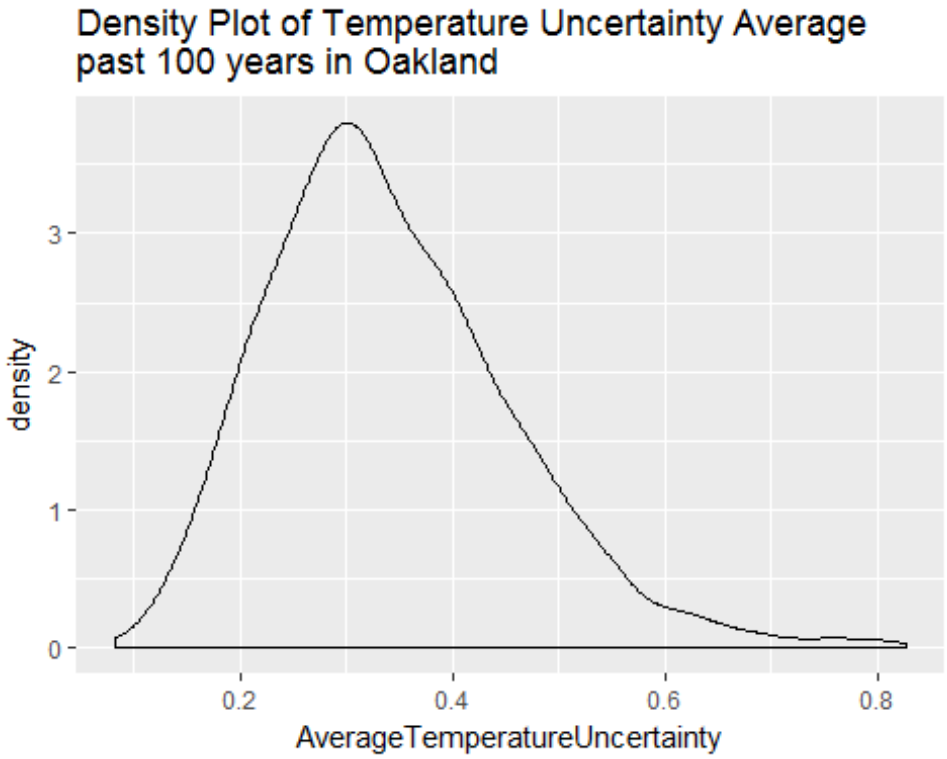
```
ggplot(oakland[Year>1916,],aes(x=dt,y=AverageTemperatureUncertainty))+
  geom_point(shape=1)+geom_smooth()+ggtitle("Temperature Uncertainty
Average\nPast 100 Years In Oakland")+
  xlab("Year")+ylab("Temperature Uncertainty Average")

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



Density Plot

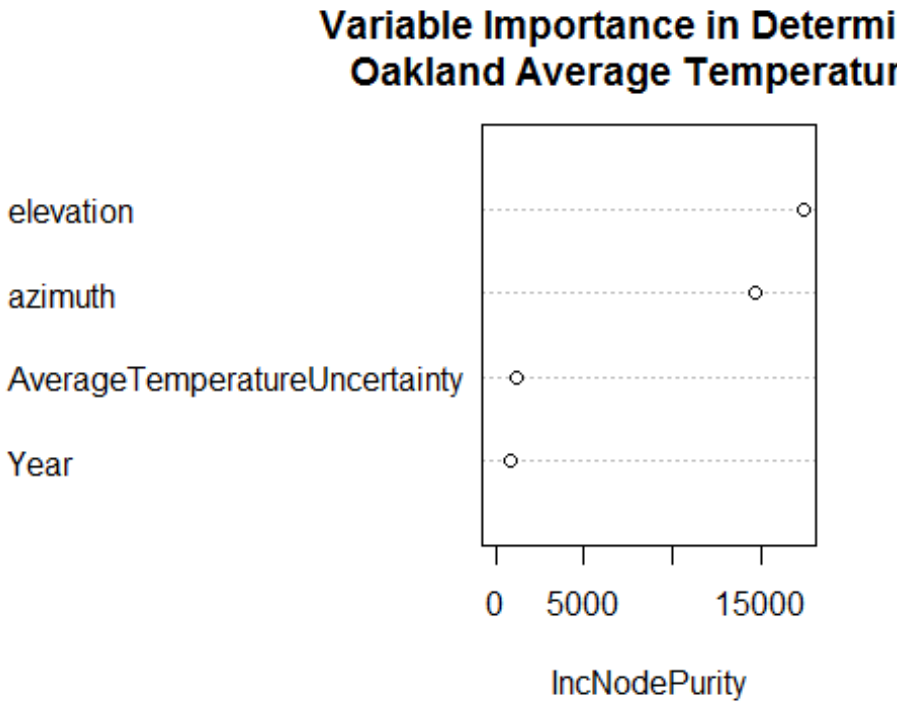
In this section, the graph demostrates the distribution on the temperature uncertainty for the last 100 years in Oakland.

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



Random Forest

```
rf<-
randomForest(subset(oakland,select=c(Year,elevation,azimuth,AverageTemp
eratureUncertainty)),oakland$AverageTemperature)
varImpPlot(rf,main="Variable Importance in Determining\nOakland Average
Temperatures")
```


Variable Importance in Determi
Oakland Average Temperatur
elevation
azimuth
AverageTemperatureUncertainty
Year
0
5000
15000
IncNodePurity