---

title: "George\_Smith\_HW6"

author: "George Smith"

date: "5/16/2021"

output: pdf\_document

---

### install

#install.packages("ggplot2")

#library("ggplot2")

###step 1: Load the Data

```{r}

air <- airquality

```

###step 2: Clean the data

```{r}

remove\_na <- function(df, n=0){

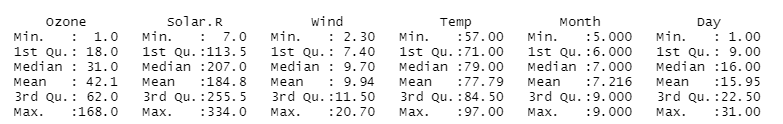
df[rowSums(is.na(df)) <= n,]

}

air2 <- remove\_na(air)

summary(air2)

```



### Step 3: Understand the data distribution

## Ozone Histogram

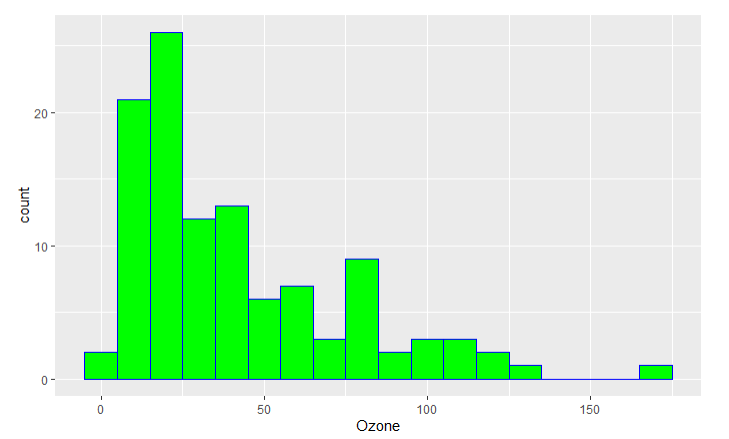
```{r}

Ozone <- ggplot(air2,aes(x=Ozone))

Ozone <- Ozone + geom\_histogram(binwidth=10,color="blue",fill="green",na.rm=TRUE)

Ozone

```



## Solar.R Histogram

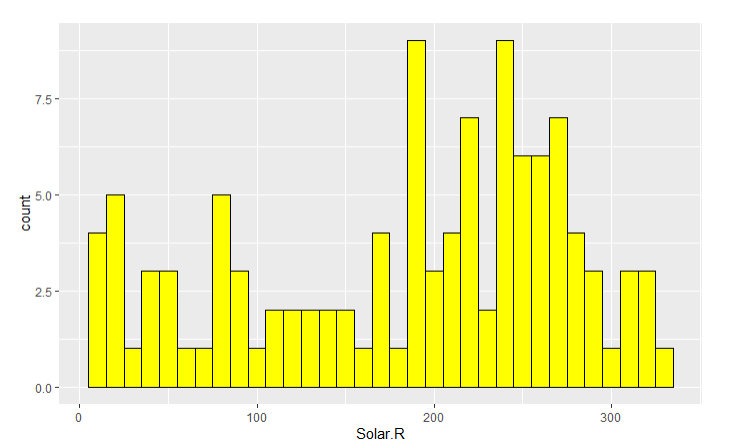
```{r}

Solar <- ggplot(air2,aes(x=Solar.R))

Solar <- Solar + geom\_histogram(binwidth=10,color="black",fill="yellow",na.rm=TRUE)

Solar

```



## Wind Historgram

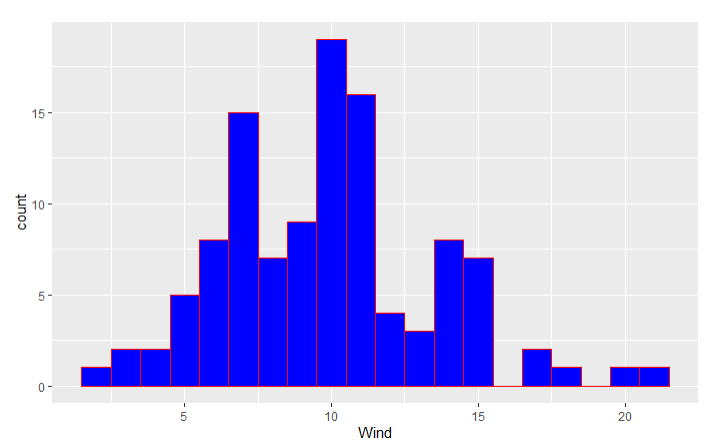
```{r}

Wind <- ggplot(air2,aes(x=Wind))

Wind <- Wind + geom\_histogram(binwidth=1,color="red",fill="blue",na.rm=TRUE)

Wind

```



## Temp Histogram

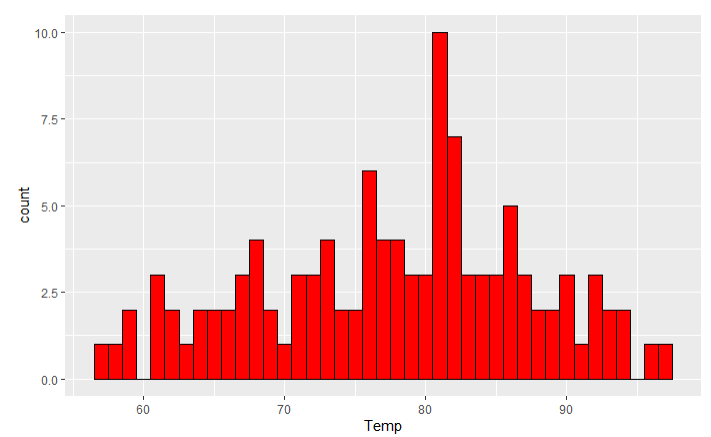
```{r}

Temp <- ggplot(air2,aes(x=Temp))

Temp <- Temp + geom\_histogram(binwidth=1,color="black",fill="red",na.rm=TRUE)

Temp

```



## Ozone boxplot

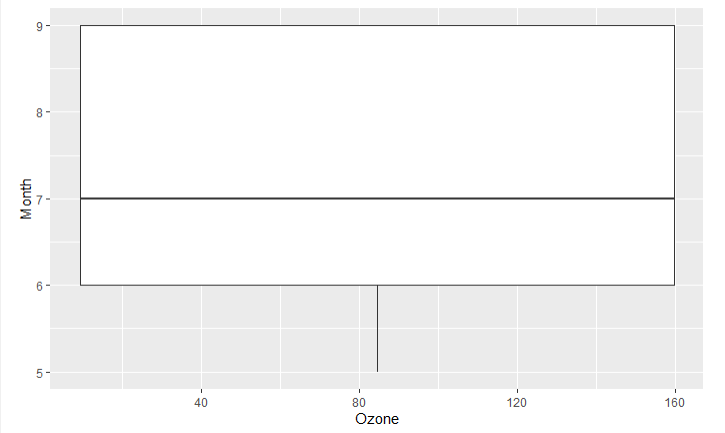
```{r}

OzoneBP <- ggplot(air2,aes(x=Ozone, y=Month))

OzoneBP <- OzoneBP + geom\_boxplot(na.rm=TRUE)

OzoneBP

```



## wind values boxplot

```{r}

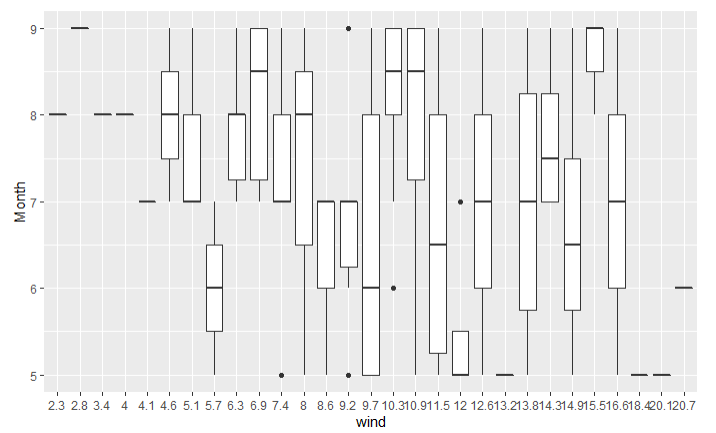
wind <- factor(air2$Wind)

windBP <- ggplot(air2, aes(x=wind, y=Month))

windBP <- windBP + geom\_boxplot(na.rm=TRUE)

windBP

```



### Explore how the data changes over time

## create data column

```{r}

date <- as.Date(with(air2, paste(1973, air2$Month, air2$Day,sep="-")), "%Y-%m-%d")

```

## Ozone date chart

```{r}

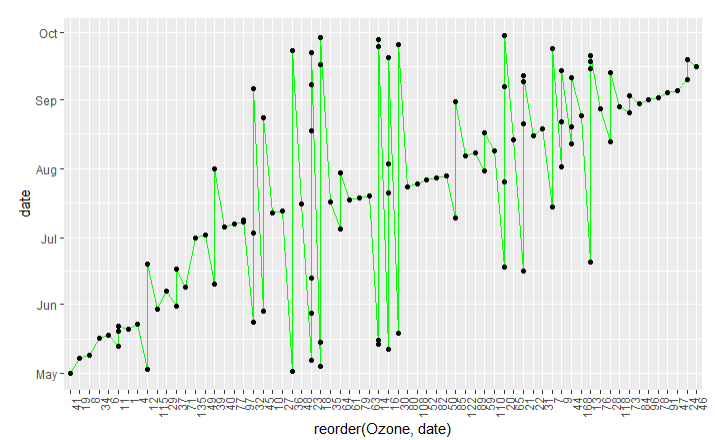
Ozone2 <- ggplot(air2, aes(x=reorder(Ozone,date),y=date,group=1))

Ozone2 <- Ozone2 + geom\_line(color="green",na.rm=TRUE)+geom\_point()

Ozone2 <- Ozone2 + theme(axis.text.x=element\_text(angle=90,hjust=1))

Ozone2

```



## temp date chart

```{r}

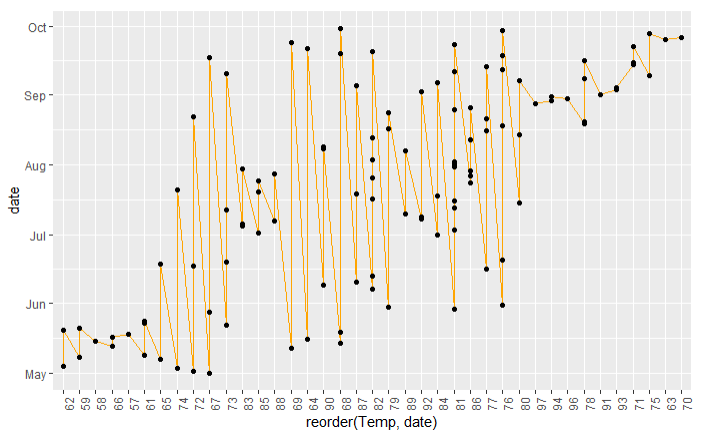
temp2 <- ggplot(air2, aes(x=reorder(Temp,date),y=date,group=1))

temp2 <- temp2 + geom\_line(color="orange",na.rm=TRUE)+geom\_point()

temp2 <- temp2 + theme(axis.text.x=element\_text(angle=90,hjust=1))

temp2

```



## Wind date chart

```{r}

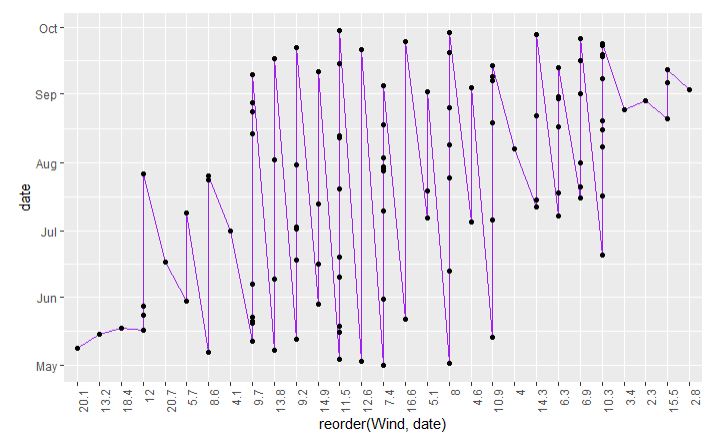
wind2 <- ggplot(air2, aes(x=reorder(Wind,date),y=date,group=1))

wind2 <- wind2 + geom\_line(color="purple",na.rm=TRUE)+geom\_point()

wind2 <- wind2 + theme(axis.text.x=element\_text(angle=90,hjust=1))

wind2

```



## Solar date chart

```{r}

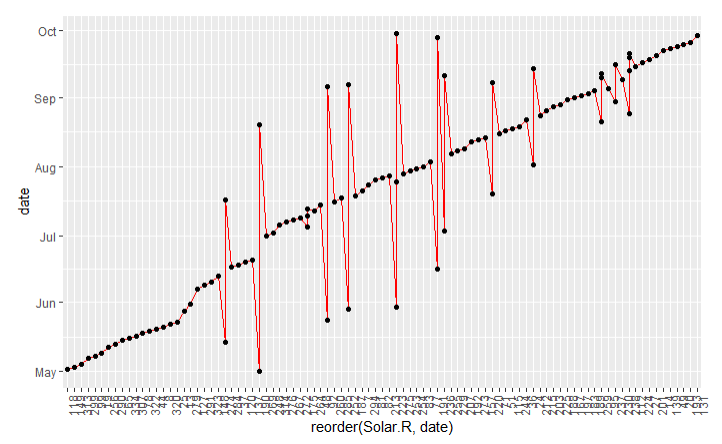
solar2 <- ggplot(air2, aes(x=reorder(Solar.R,date),y=date,group=1))

solar2 <- solar2 + geom\_line(color="red",na.rm=TRUE)+geom\_point()

solar2 <- solar2 + theme(axis.text.x=element\_text(angle=90,hjust=1))

solar2

```



## one chart with 4 lines

```{r}

plot <- ggplot(air2,aes(x=date),group=1)

plot <- plot + geom\_line(aes(y=Ozone,color="Ozone"),na.rm=TRUE)

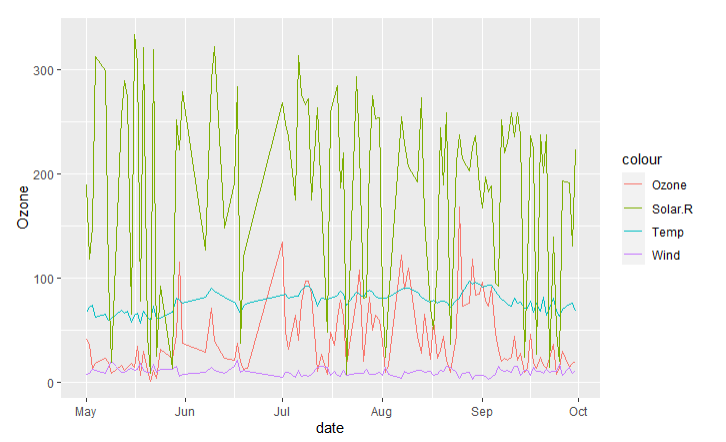
plot <- plot + geom\_line(aes(y=Temp,color="Temp"),na.rm=TRUE)

plot <- plot + geom\_line(aes(y=Wind,color="Wind"),na.rm=TRUE)

plot <- plot + geom\_line(aes(y=Solar.R,color="Solar.R"),na.rm=TRUE)

plot

```



## Step 4: Look at all the data via Heatmap

```{r}

heat<- ggplot(air2, aes(x=Day,group=Day))

heat<- heat + geom\_tile(aes(y = Ozone), color="red",na.rm=TRUE)

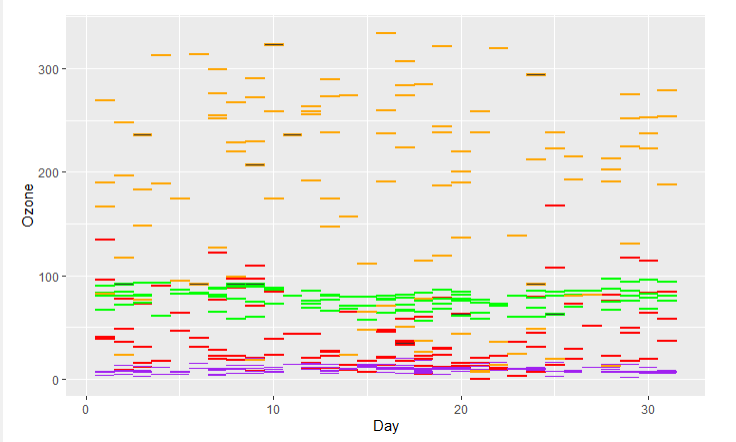
heat<- heat + geom\_tile(aes(y = Temp), color="green",na.rm=TRUE)

heat<- heat + geom\_tile(aes(y = Wind), color="purple",na.rm=TRUE)

heat<- heat + geom\_tile(aes(y = Solar.R), color="orange",na.rm=TRUE)

heat

```



### step 5 Look at all the data via a scatter chart

```{r}

scatter <- ggplot(air2,aes(x=Wind,y=Temp),na.rm=TRUE)

scatter <- scatter+geom\_point(aes(size=Ozone,color=Solar.R),na.rm=TRUE)

scatter

```

### Step 6 Final Analysis

## Do you see any patterns after expoloring the data

# yes all variables appear to have a positive linear trend with time, which appears to support global warming.

## What was the most useful visualization

# the most useful visulization for me are the line charts as they clearly depict trends related to time.