

# New York Air Quality

Han Wang

17/7/2017

## A.Basic Plotting

### 1.Distribution of measurements

```
data("airquality")  
require(ggplot2)
```

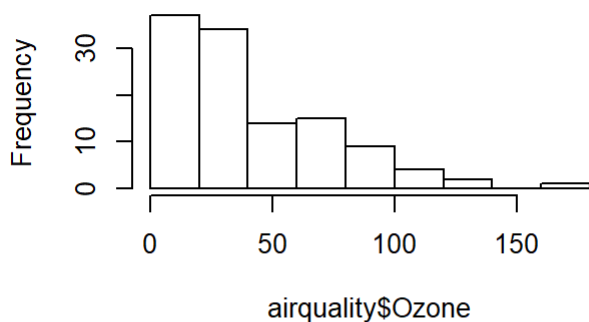
```
## Loading required package: ggplot2
```

```
require(lattice)
```

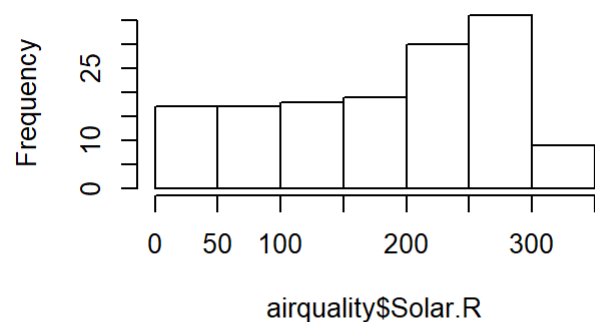
```
## Loading required package: lattice
```

```
par(mfrow=c(2,2))  
hist(airquality$Ozone)  
hist(airquality$Solar.R)  
hist(airquality$Wind)  
hist(airquality$Temp)
```

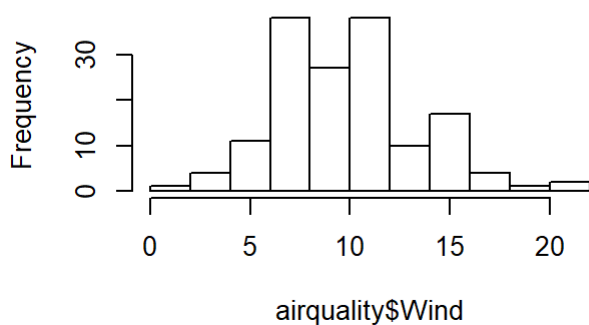
**Histogram of airquality\$Ozone**



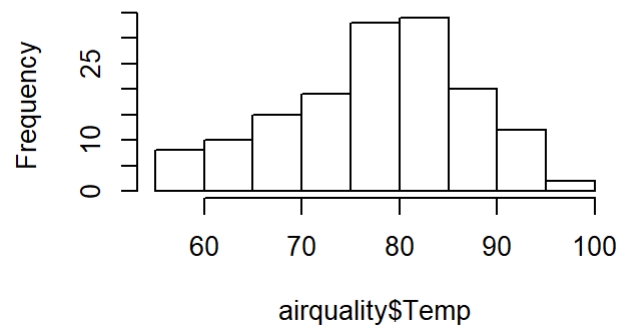
**Histogram of airquality\$Solar.R**



**Histogram of airquality\$Wind**



**Histogram of airquality\$Temp**

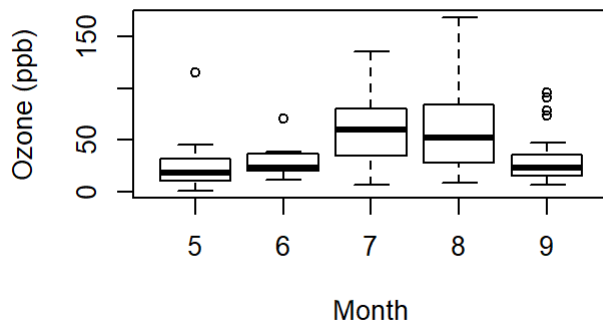
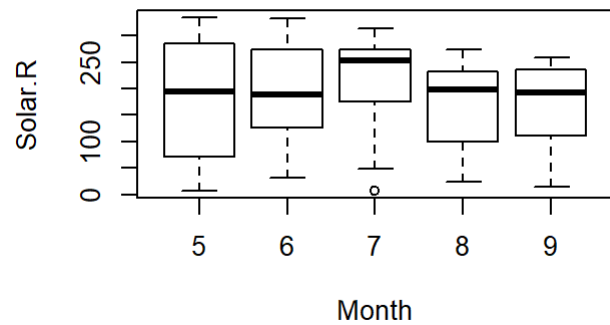
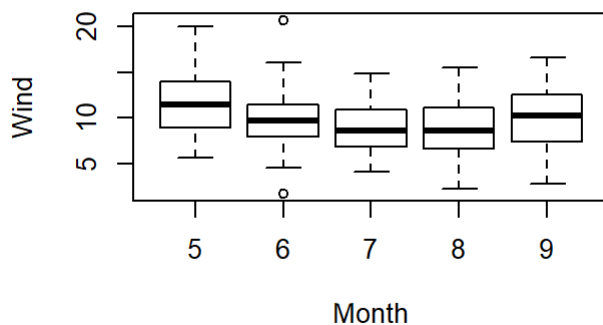
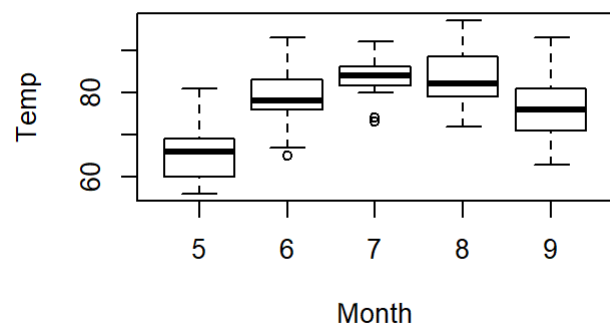


### 2.Boxplot of measurements

```

airquality <- transform(airquality, Month = factor(Month))
par(mfrow=c(2,2))
boxplot(Ozone ~ Month, airquality, xlab = "Month", ylab = "Ozone (ppb)", main="Boxplot of Ozone by Month")
boxplot(Solar.R ~ Month, airquality, xlab = "Month", ylab = "Solar.R", main="Boxplot of Solar.R by Month")
boxplot(Wind ~ Month, airquality, xlab = "Month", ylab = "Wind", main="Boxplot of Wind by Month")
boxplot(Temp ~ Month, airquality, xlab = "Month", ylab = "Temp", main="Boxplot of Temp by Month")

```

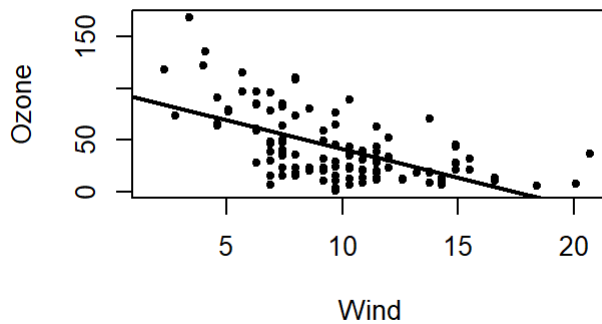
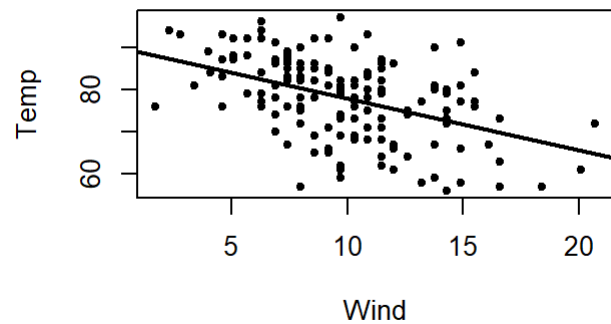
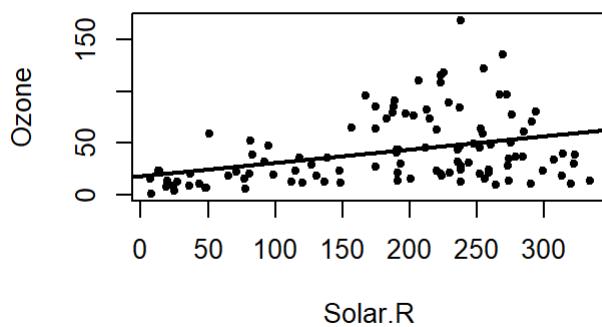
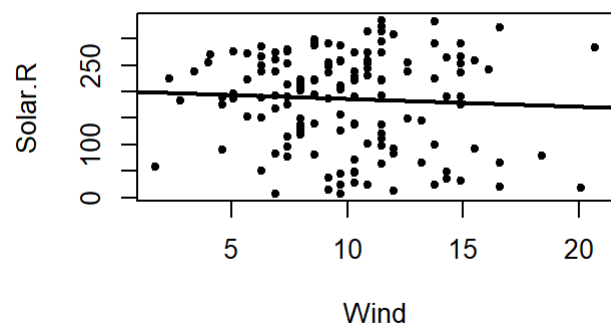
**Boxplot of Ozone by Month****Boxplot of Solar.R by Month****Boxplot of Wind by Month****Boxplot of Temp by Month**

### 3. Correlation-ScatterPlot

```

par(mfrow=c(2,2))
with(airquality, plot(Wind, Ozone, main = "Ozone and Wind in New York City", pch = 20))
model <- lm(Ozone ~ Wind, airquality)
abline(model, lwd = 2)
with(airquality, plot(Wind, Temp, main = "Temp and Wind in New York City", pch = 20))
model <- lm(Temp ~ Wind, airquality)
abline(model, lwd = 2)
with(airquality, plot(Solar.R, Ozone, main = "Ozone and Solar.R in New York City", pch = 20))
model <- lm(Ozone ~ Solar.R, airquality)
abline(model, lwd = 2)
with(airquality, plot(Wind, Solar.R, main = "Solar.R and Wind in New York City", pch = 20))
model <- lm(Solar.R ~ Wind, airquality)
abline(model, lwd = 2)

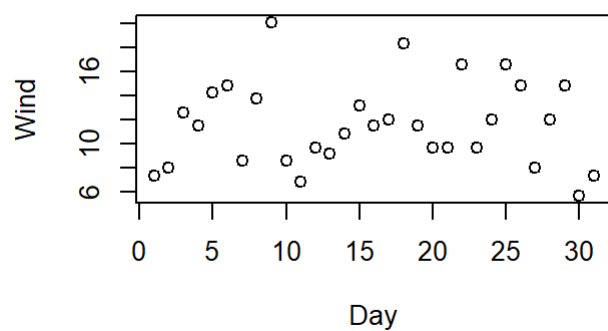
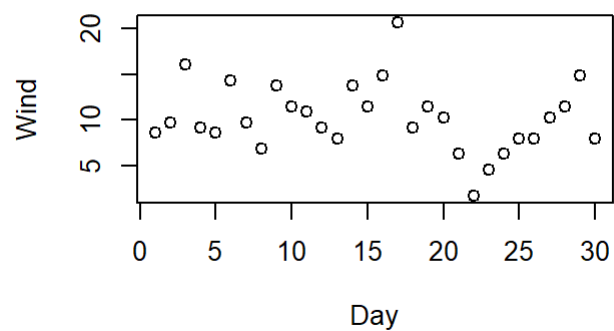
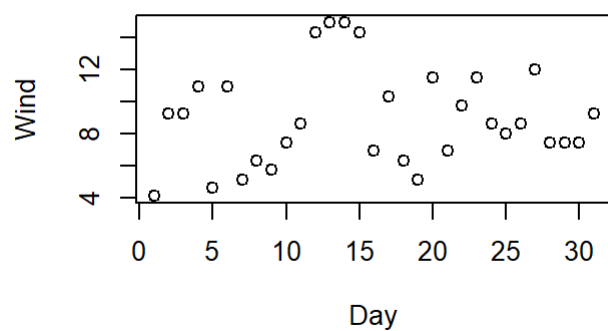
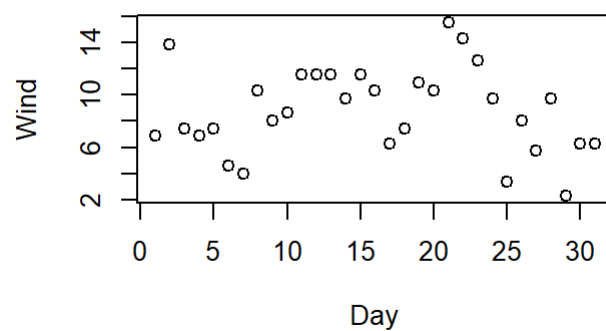
```

**Ozone and Wind in New York City****Temp and Wind in New York City****Ozone and Solar.R in New York City****Solar.R and Wind in New York City**

#### 4. Airquality by Month

```
May<-subset(airquality,airquality$Month=='5')
June<-subset(airquality,airquality$Month=='6')
July<-subset(airquality,airquality$Month=='7')
August<-subset(airquality,airquality$Month=='8')
September<-subset(airquality,airquality$Month=='9')
par(mfrow = c(2, 2))

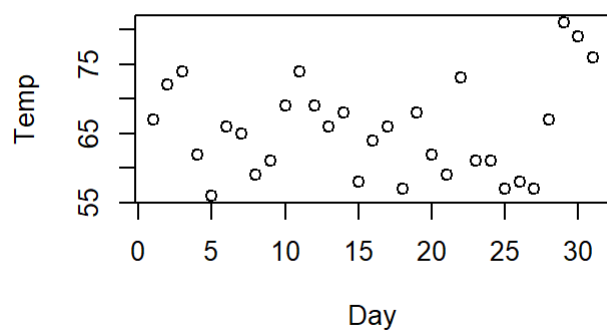
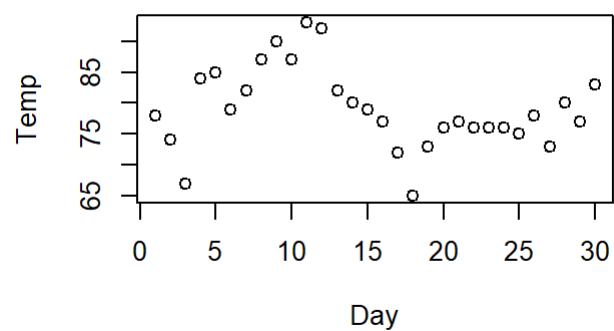
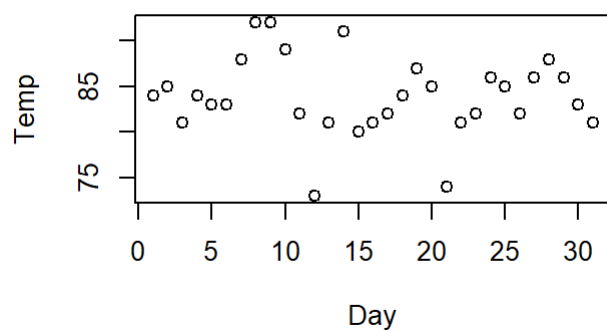
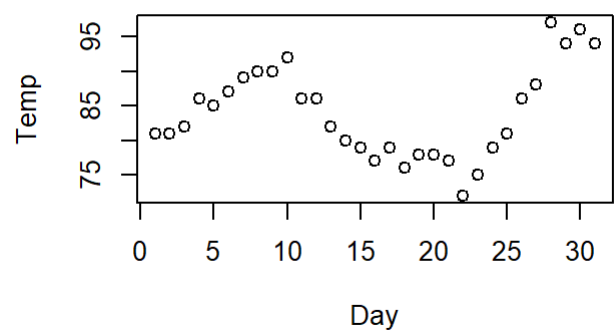
with(May,plot(Day,Wind,main="Wind in May"),type = "l", pch = 20)+with(June,plot(Day,Wind,main="Wind in
June"),type = "l", pch = 20)+with(July,plot(Day,Wind,main="Wind in July"),type = "l", pch =
20)+with(August,plot(Day,Wind,main="Wind in August"),type = "l", pch = 20)
```

**Wind in May****Wind in June****Wind in July****Wind in August**

```
## numeric(0)
```

```
par(mfrow = c(2, 2))
```

```
with(May, plot(Day, Temp, main="Temperature in May"), pch = 20)+with(June, plot(Day, Temp, main="Temperature
  in June"), pch = 20)+with(July, plot(Day, Temp, main="Temperature in July"), pch = 20)+with(August, plot(Da
    y, Temp, main="Temperature in August"), pch = 20)
```

**Temperature in May****Temperature in June****Temperature in July****Temperature in August**

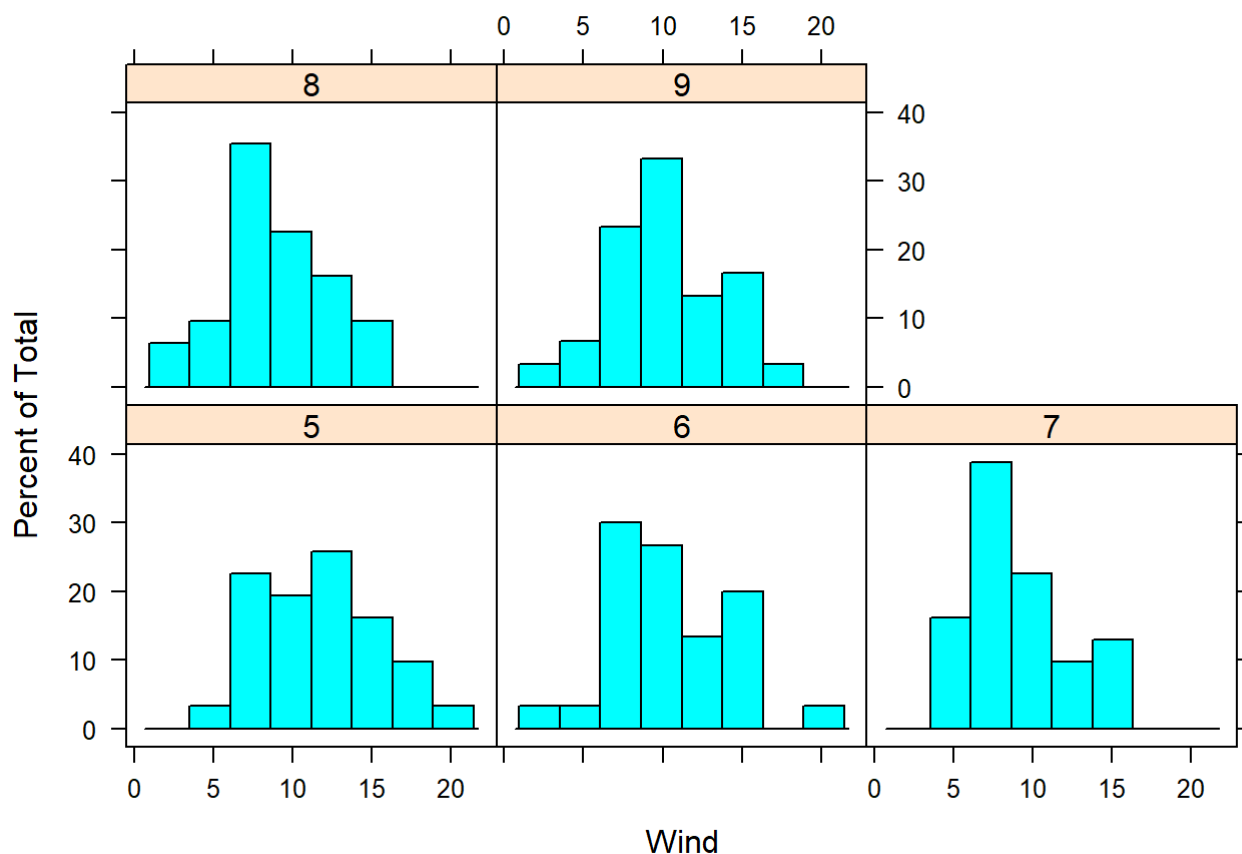
```
## numeric(0)
```

## B.Lattice

### 1. Histogram of measurements

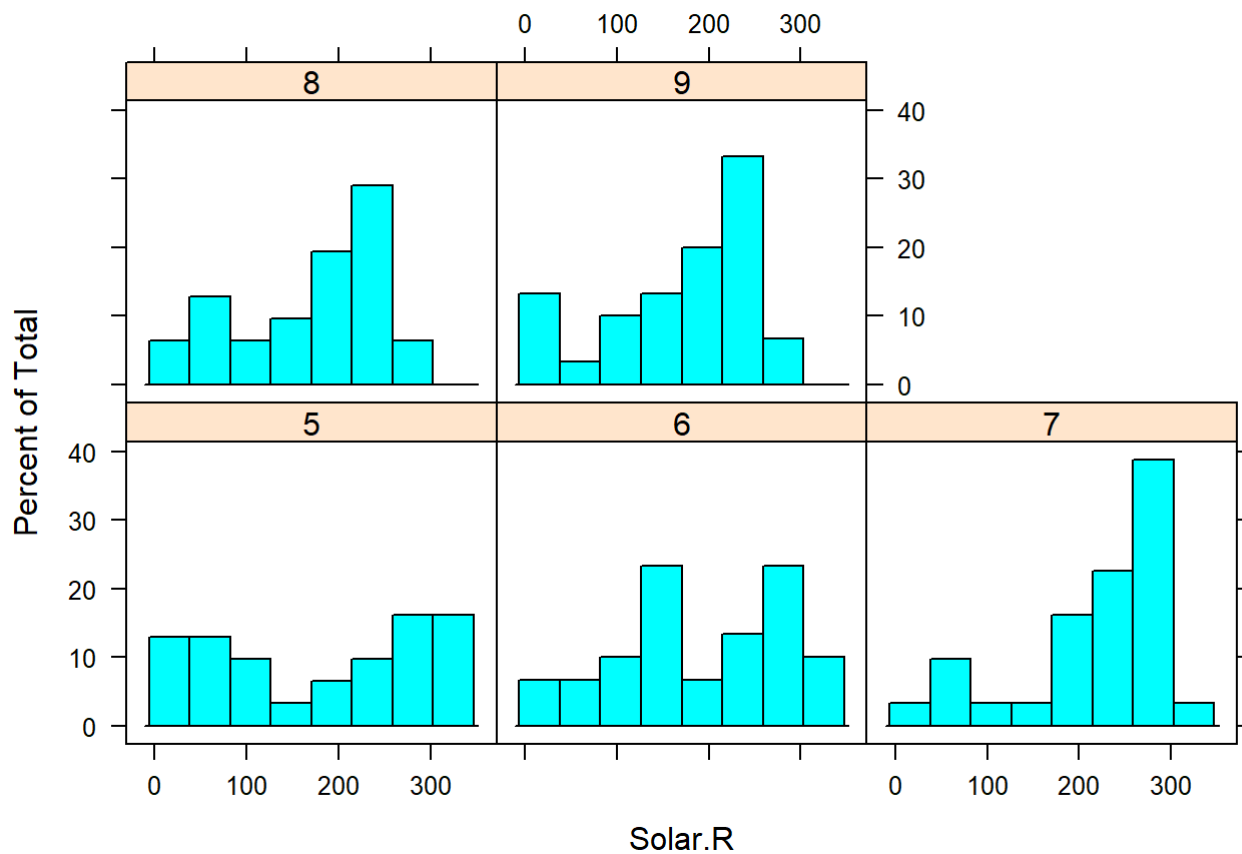
```
histogram(~ Wind | Month, data=airquality, main="Wind by month")
```

## Wind by month

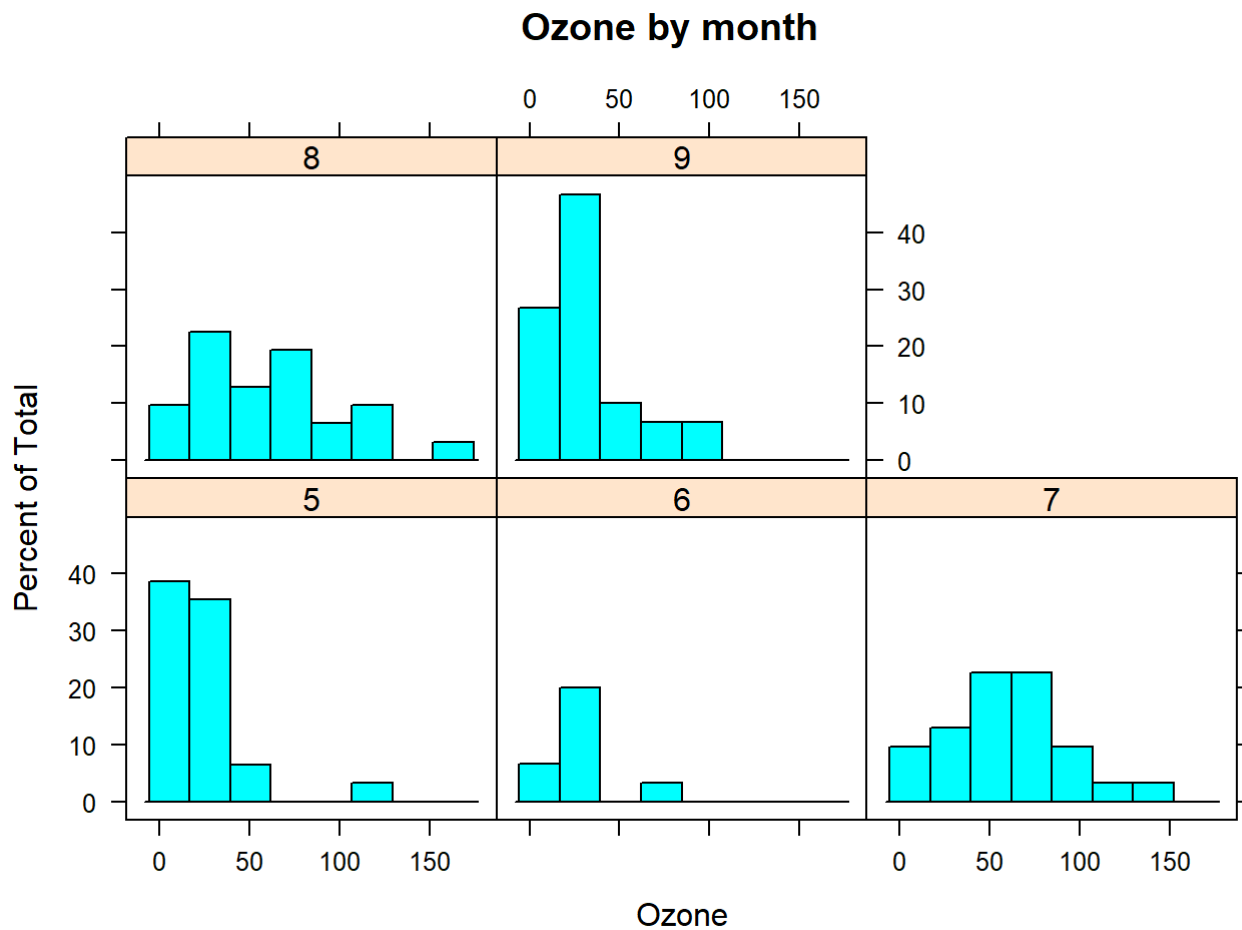


```
histogram( ~ Solar.R | Month, data=airquality, main="Solar.R by month")
```

## Solar.R by month

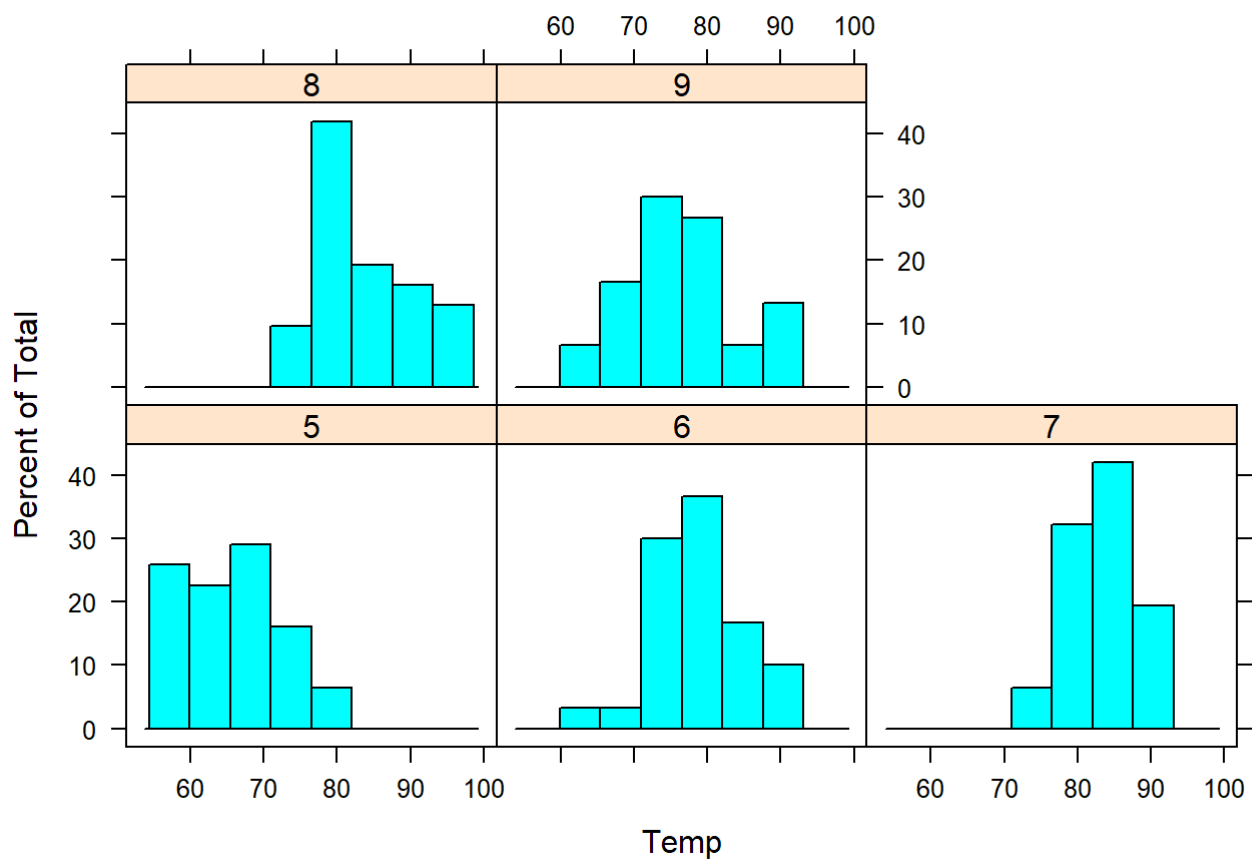


```
histogram( ~ Ozone | Month, data=airquality, main="Ozone by month")
```



```
histogram( ~ Temp | Month, data=airquality, main="Temperature by month")
```

## Temperature by month

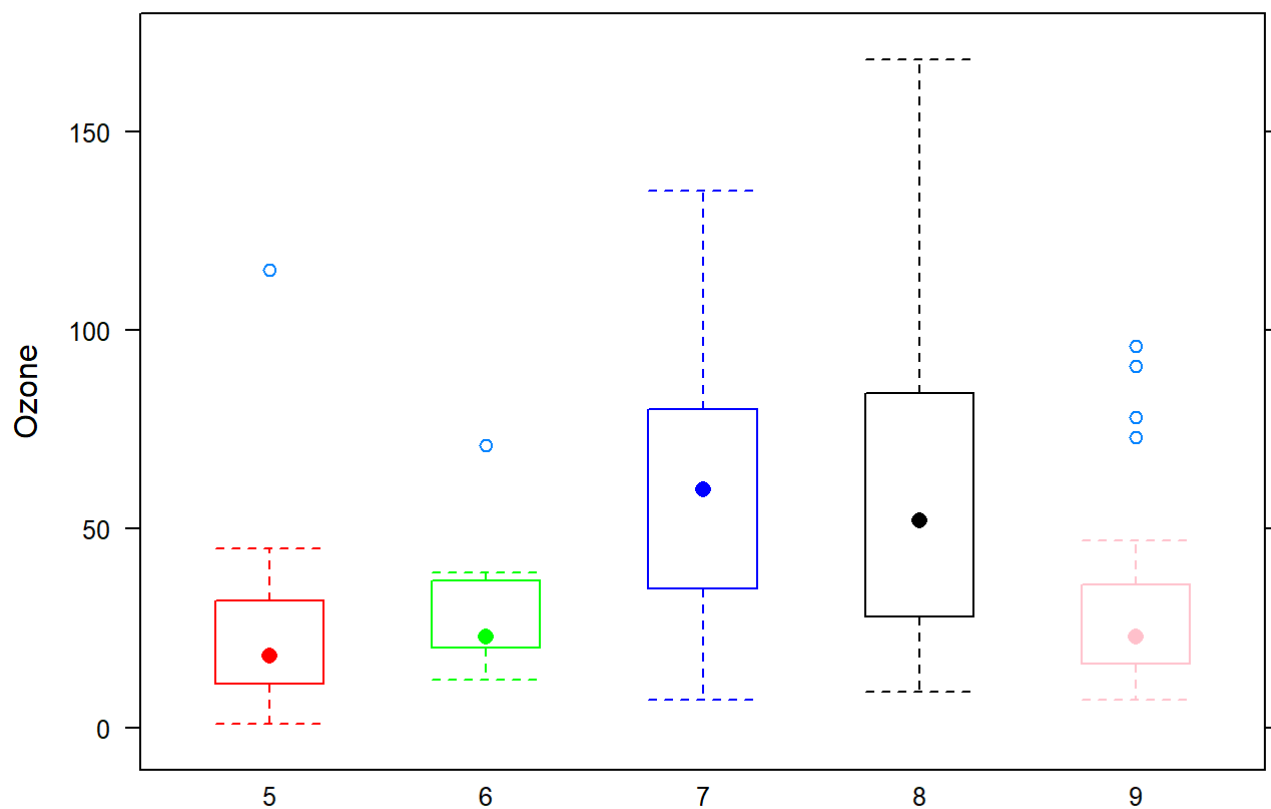


## 2. Boxplots by month

```
bwplot( Ozone ~ Month, data=airquality, main="Ozone by month", par.settings = list(box.umbrella=list(col=
c("red", "green", "blue", "black", "pink")), box.dot=list(col= c("red", "green",
"blue", "black", "pink")), box.rectangle = list(col= c("red", "green", "blue", "black", "pink"))))
```

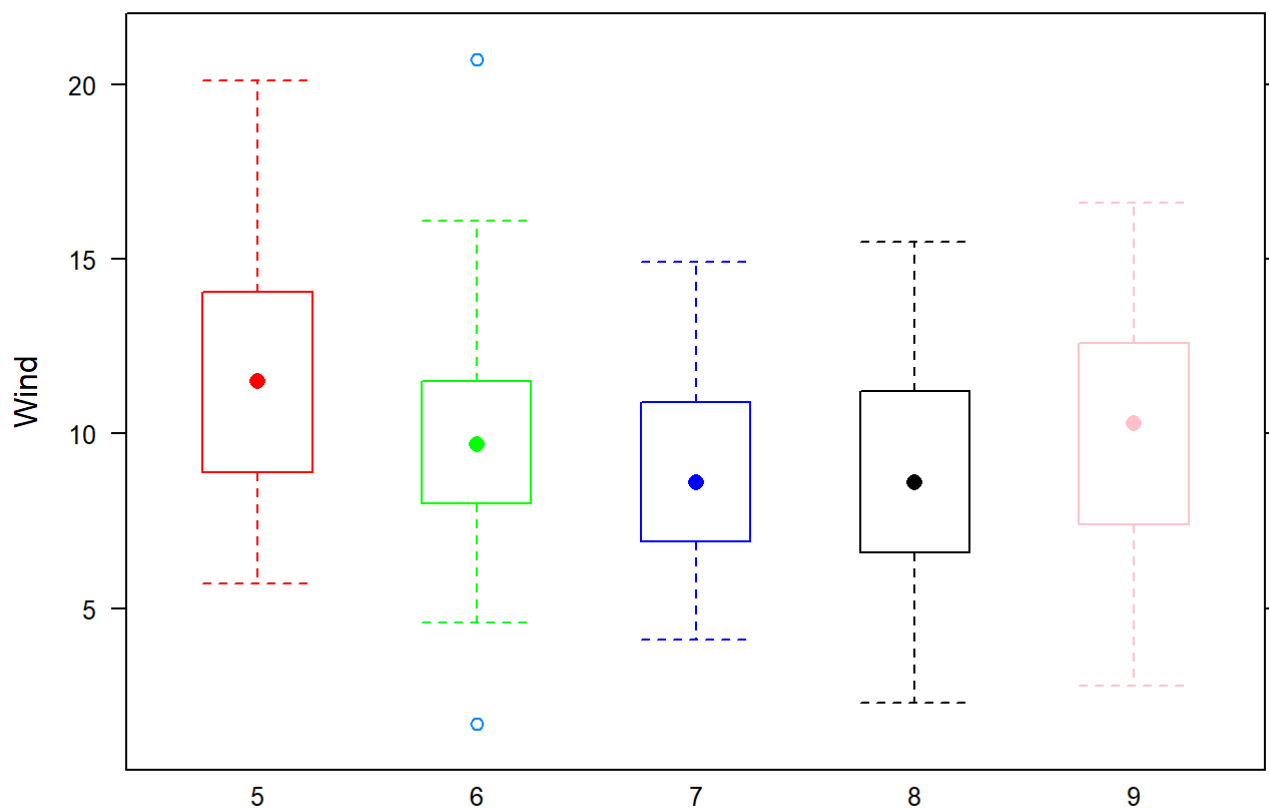


## Ozone by month



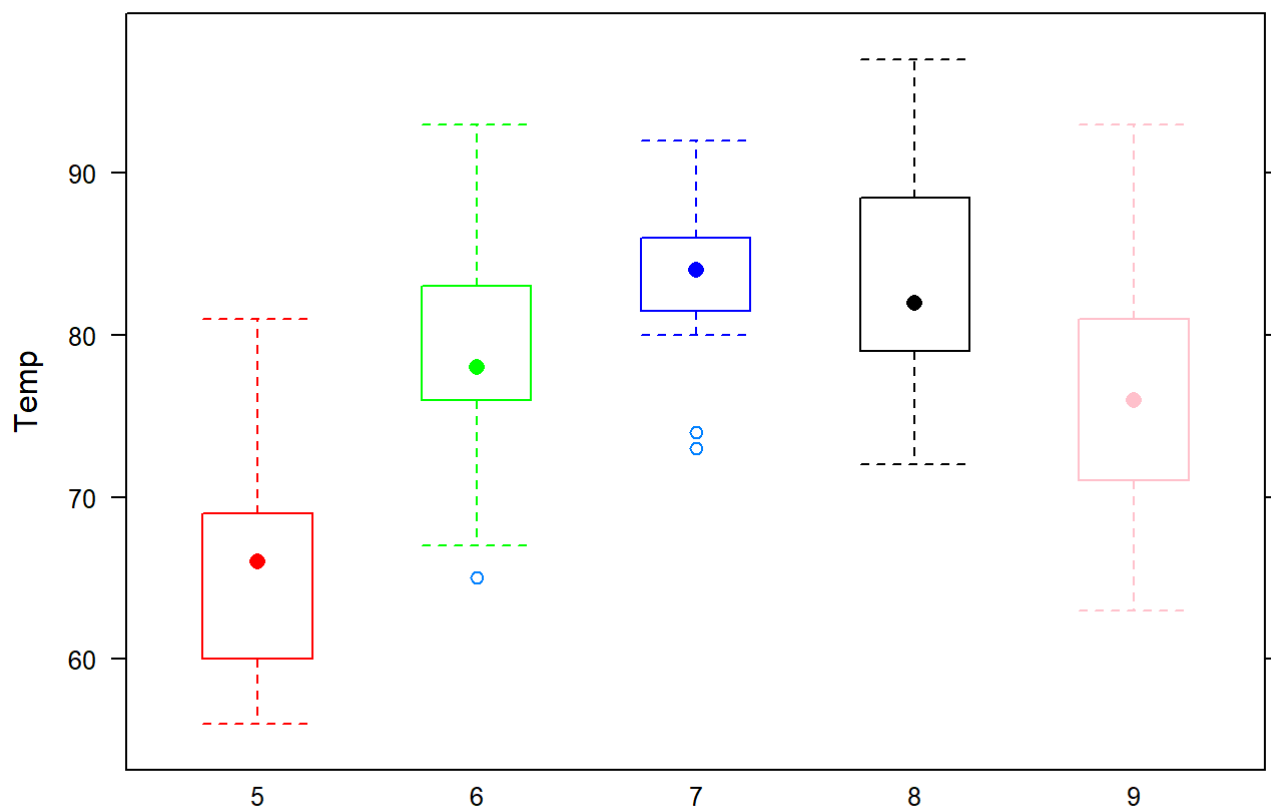
```
bwplot( Wind ~Month, data=airquality,main="Wind by month",par.settings = list(box.umbrella=list(col=
c("red", "green", "blue","black","pink")), box.dot=list(col= c("red", "green",
"blue","black","pink")), box.rectangle = list(col= c("red", "green", "blue","black","pink"))))
```

## Wind by month



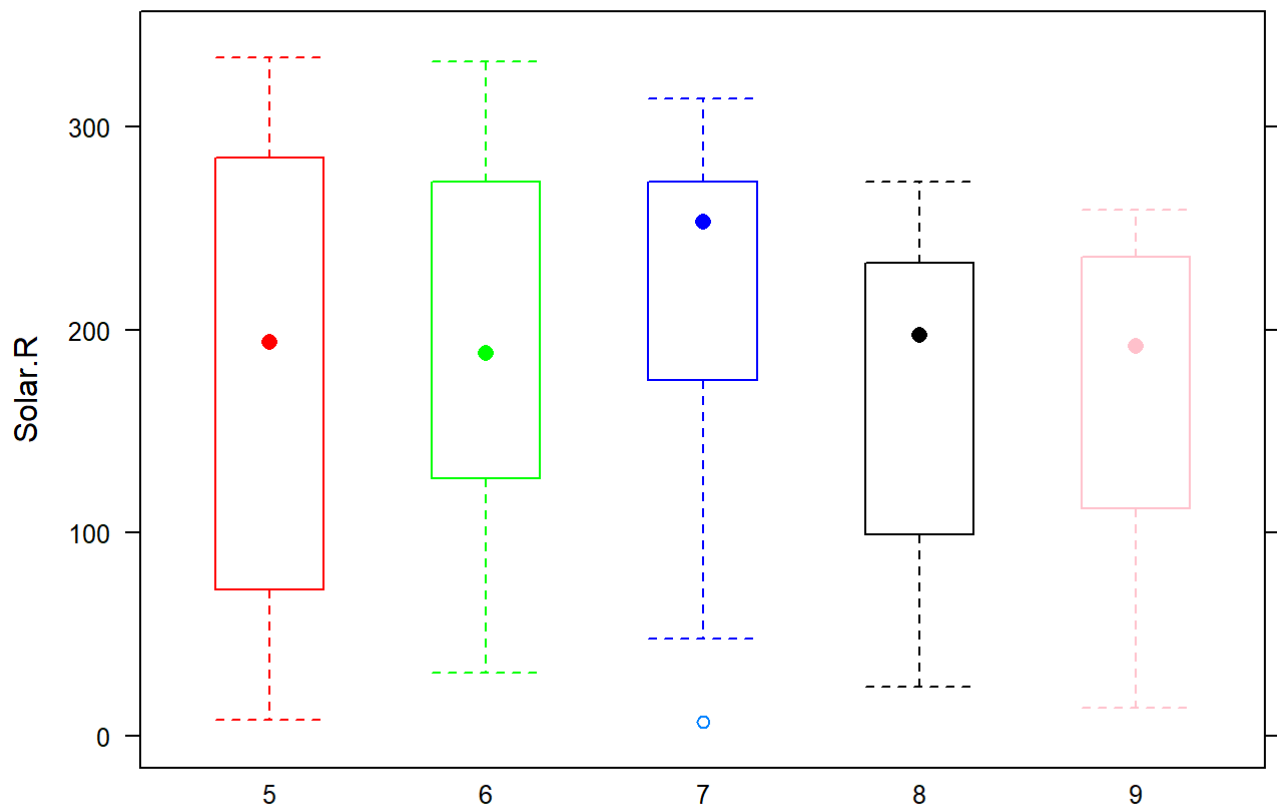
```
bwplot( Temp ~Month, data=airquality,main="Temperature by month",par.settings =
list(box.umbrella=list(col= c("red", "green", "blue","black","pink")), box.dot=list(col= c("red", "gre
en", "blue","black","pink")), box.rectangle = list(col= c("red", "green", "blue","black","pink"))))
```

## Temperature by month



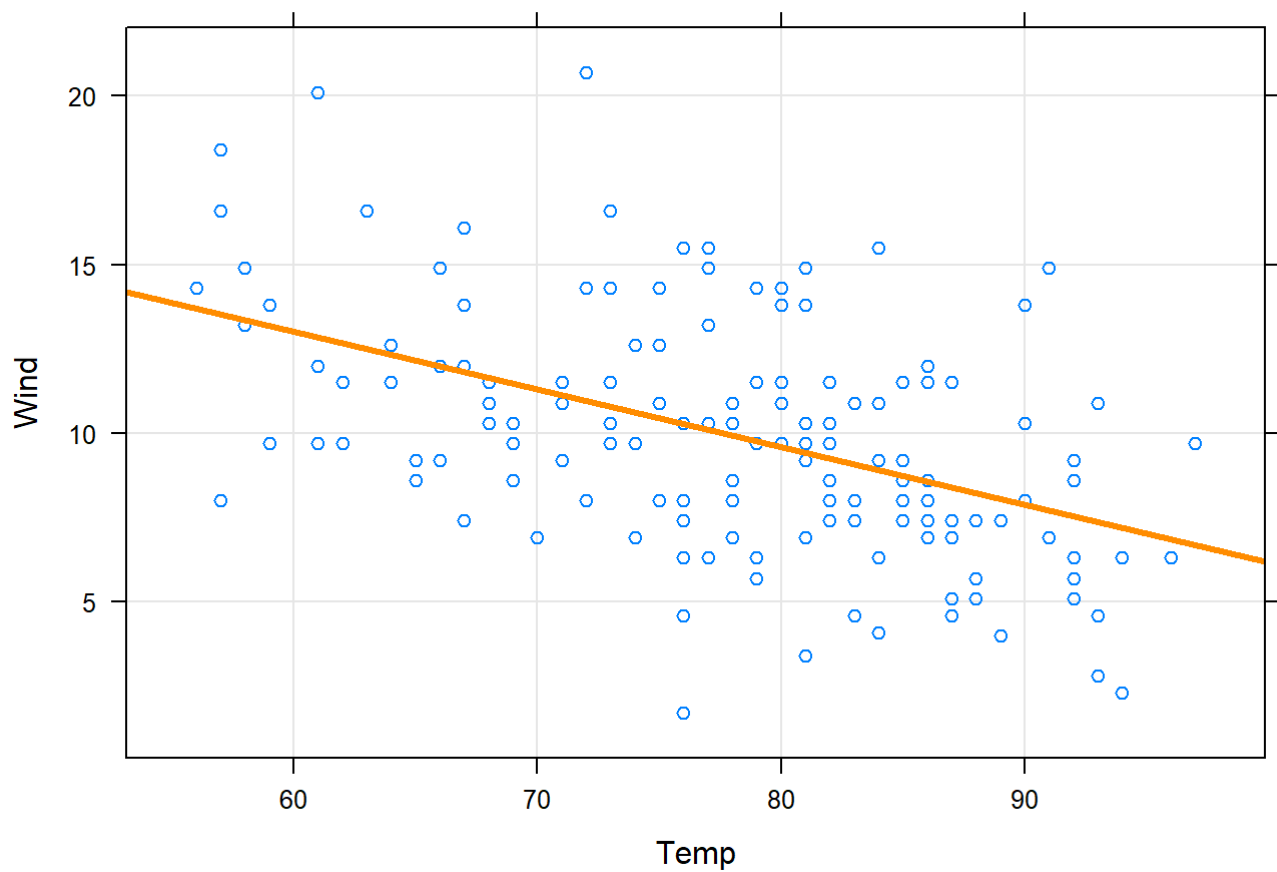
```
bwplot( Solar.R ~Month, data=airquality,main="Solar.R by month",par.settings =
list(box.umbrella=list(col= c("red", "green", "blue","black","pink")), box.dot=list(col= c("red", "gre
en", "blue","black","pink")), box.rectangle = list(col= c("red", "green", "blue","black","pink"))))
```

## Solar.R by month

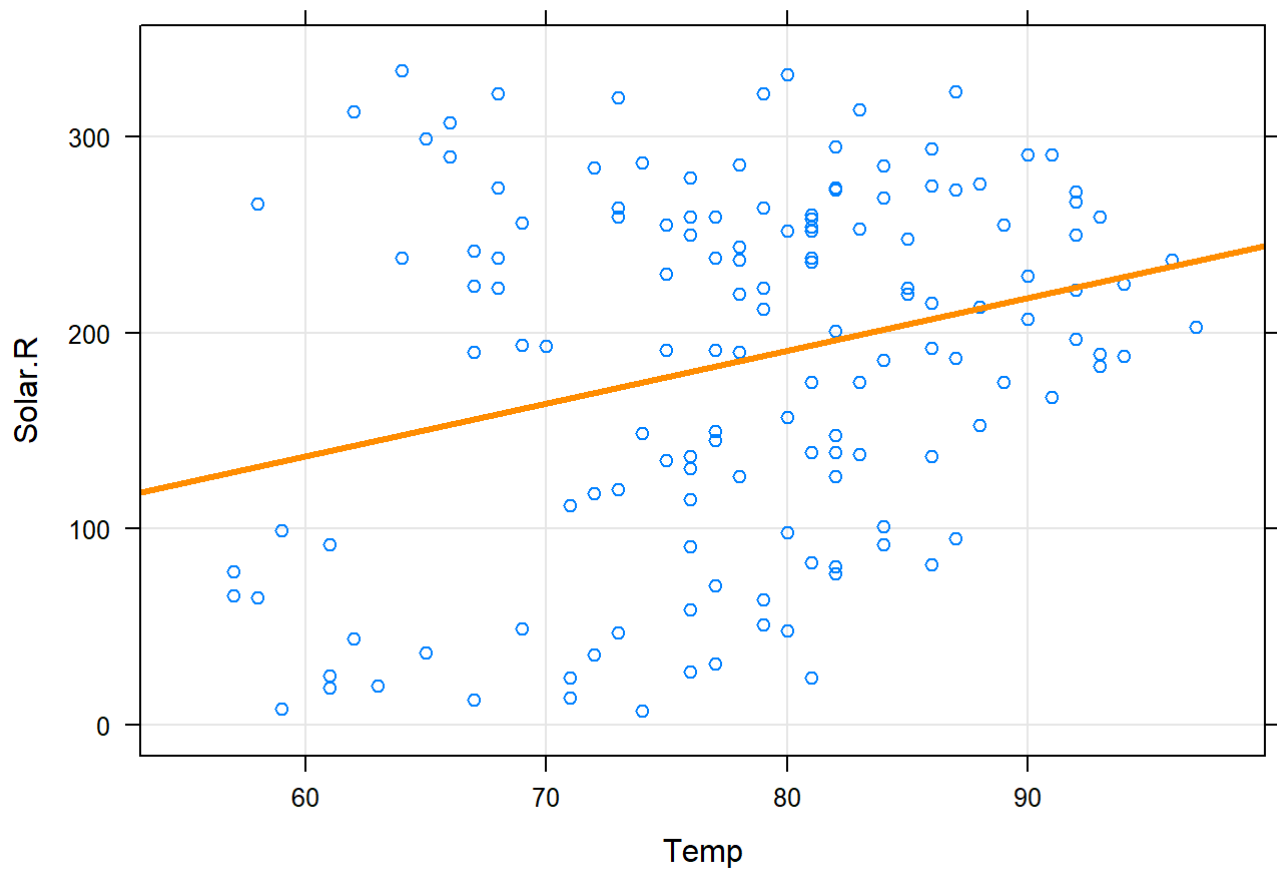


### 3. Correlation-Scatter plots

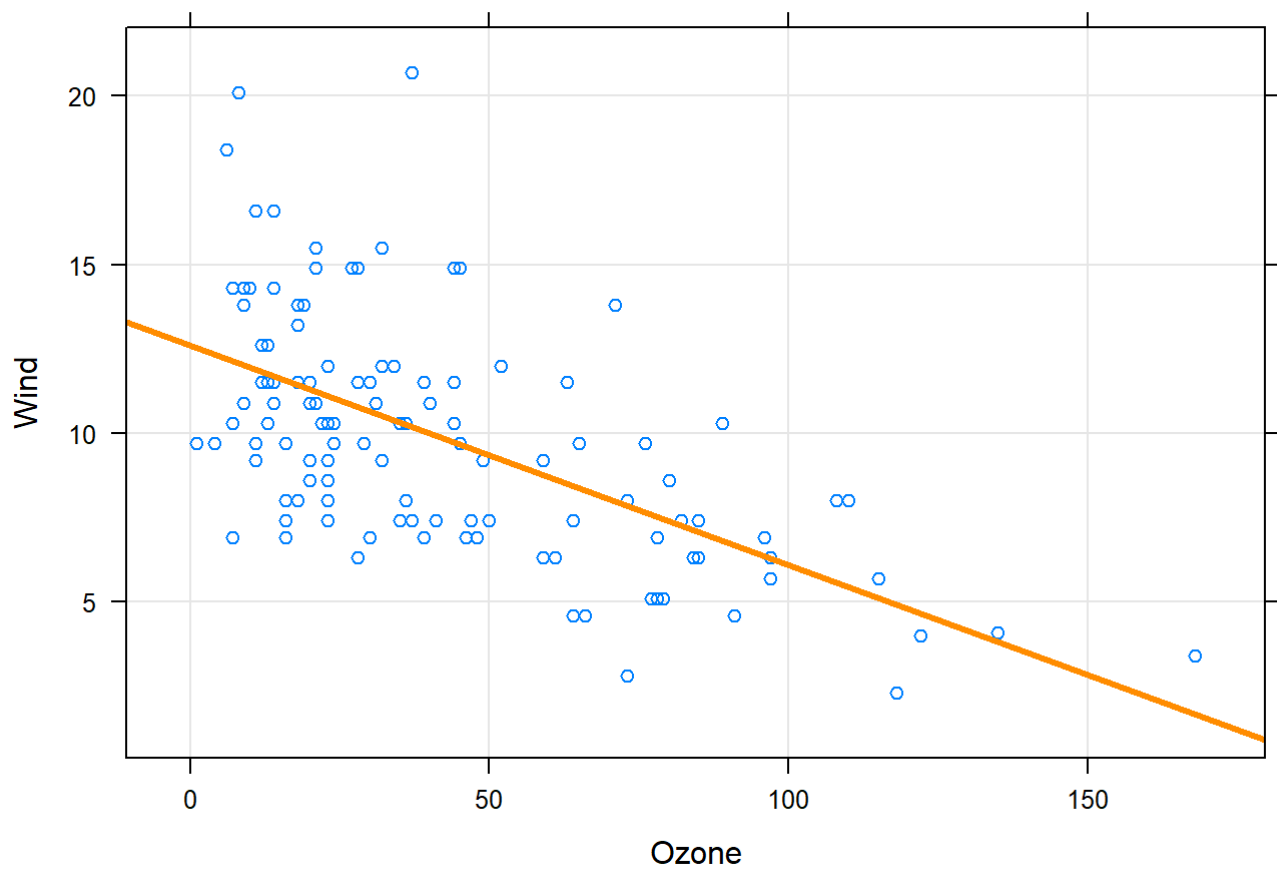
```
xyplot(Wind ~ Temp, airquality,  
       grid = TRUE,  
       type = c("p", "r"), col.line = "darkorange", lwd = 3)
```



```
xyplot(Solar.R ~ Temp, airquality,  
       grid = TRUE,  
       type = c("p", "r"), col.line = "darkorange", lwd = 3)
```



```
xyplot(Wind ~ Ozone, airquality,  
       grid = TRUE,  
       type = c("p", "r"), col.line = "darkorange", lwd = 3)
```

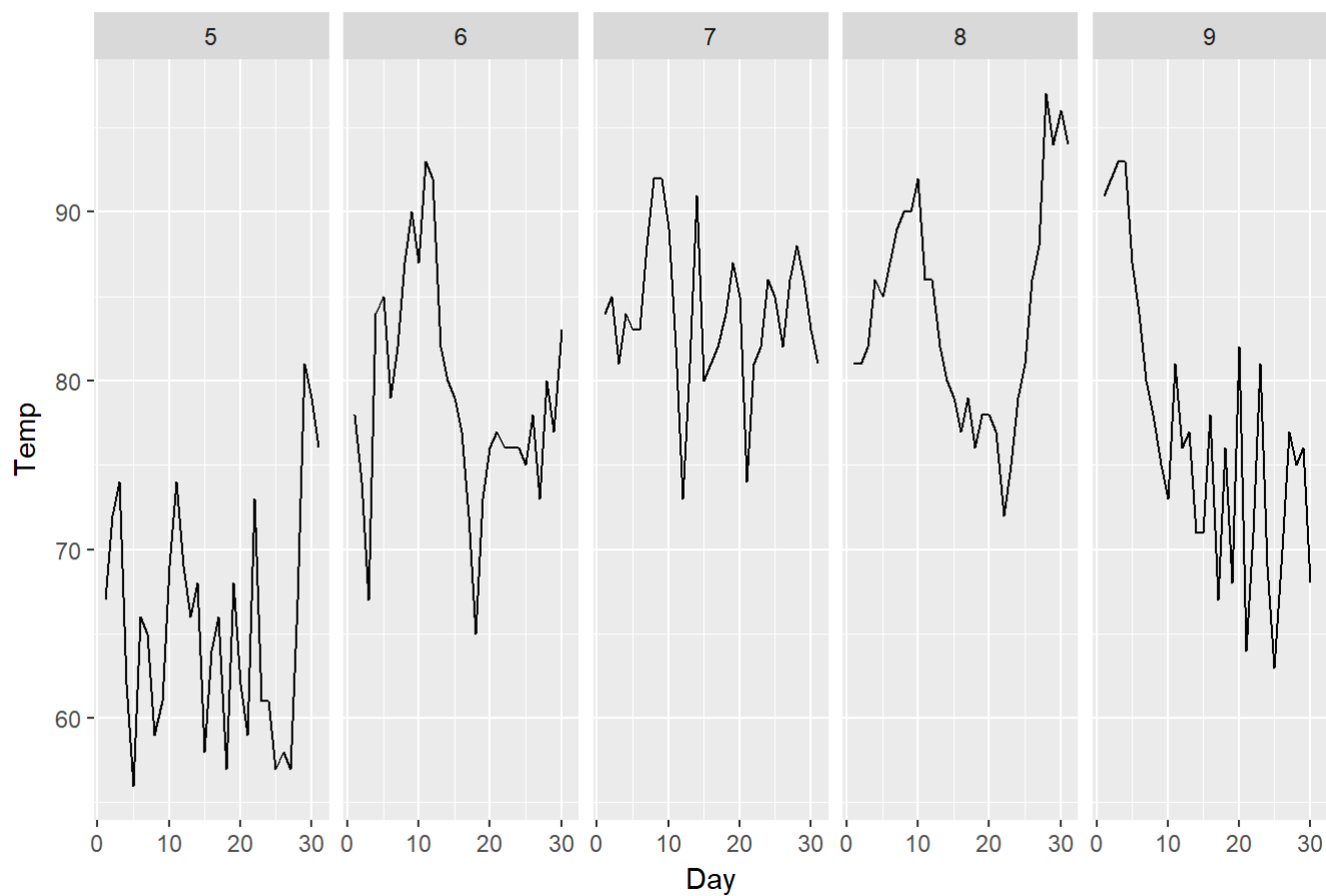


## C.ggplot2

### 1. Temperature from May to September

```
g<-ggplot(airquality, aes(x = Day, y = Temp))  
g+ geom_line() +facet_grid(. ~ Month)+ggtitle("Temperature from May to September")
```

## Temperature from May to September

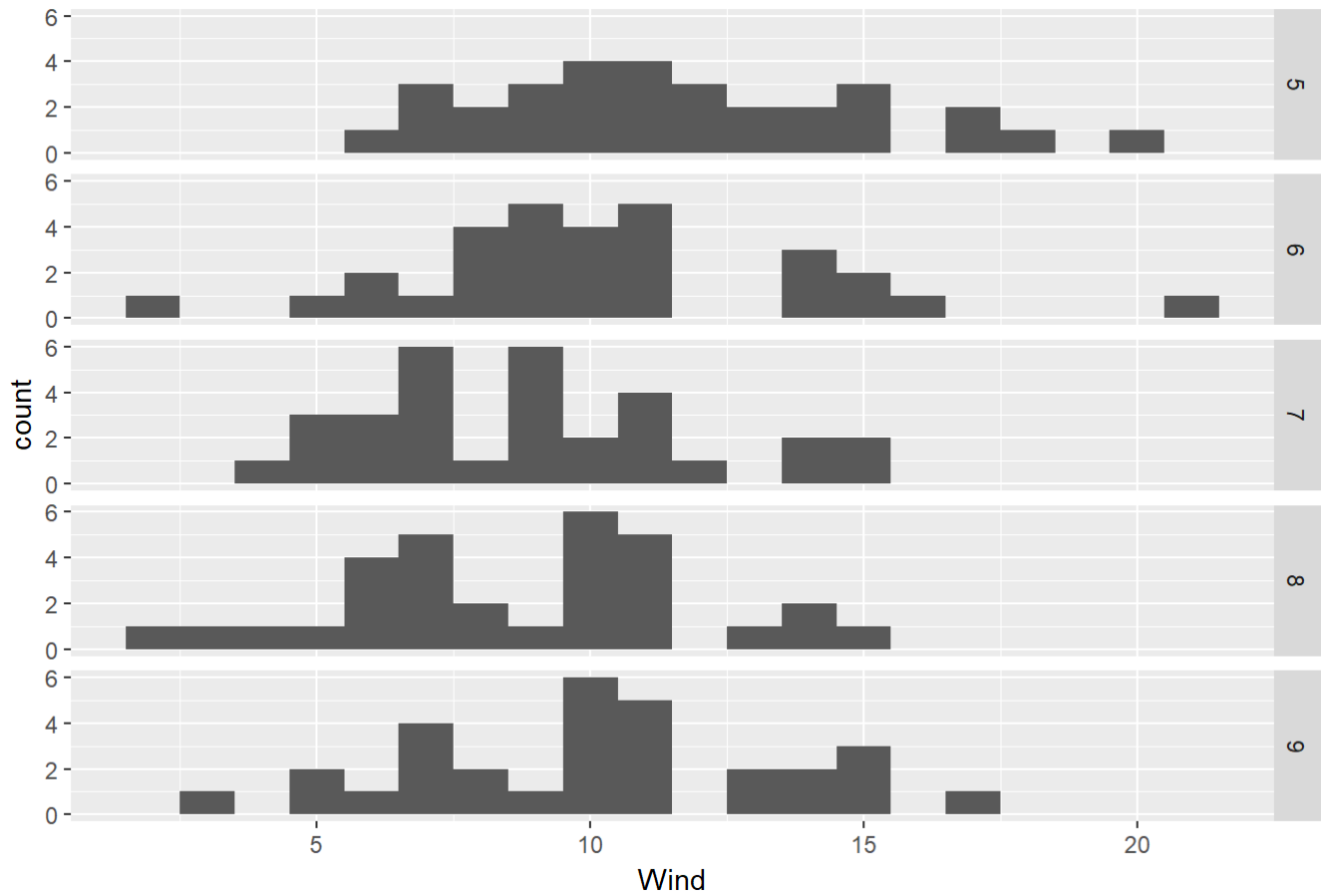


## 2. Distribution of measurement by month

```
qplot(Wind, data = airquality, facets = Month ~ ., binwidth = 1, main="Wind by month")
```

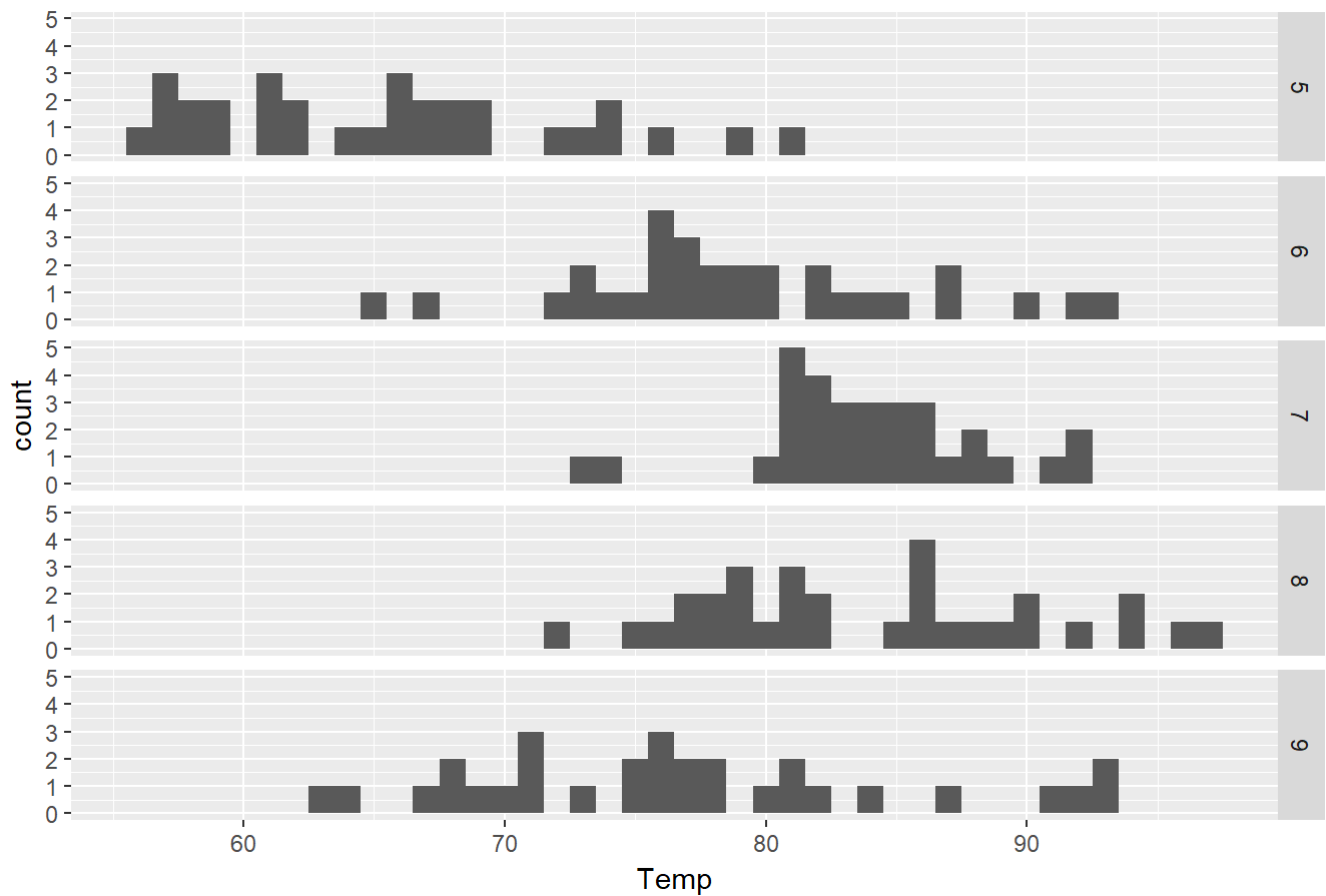


## Wind by month



```
qplot(Temp, data = airquality, facets = Month ~ ., binwidth = 1, main="Temperature by month")
```

## Temperature by month



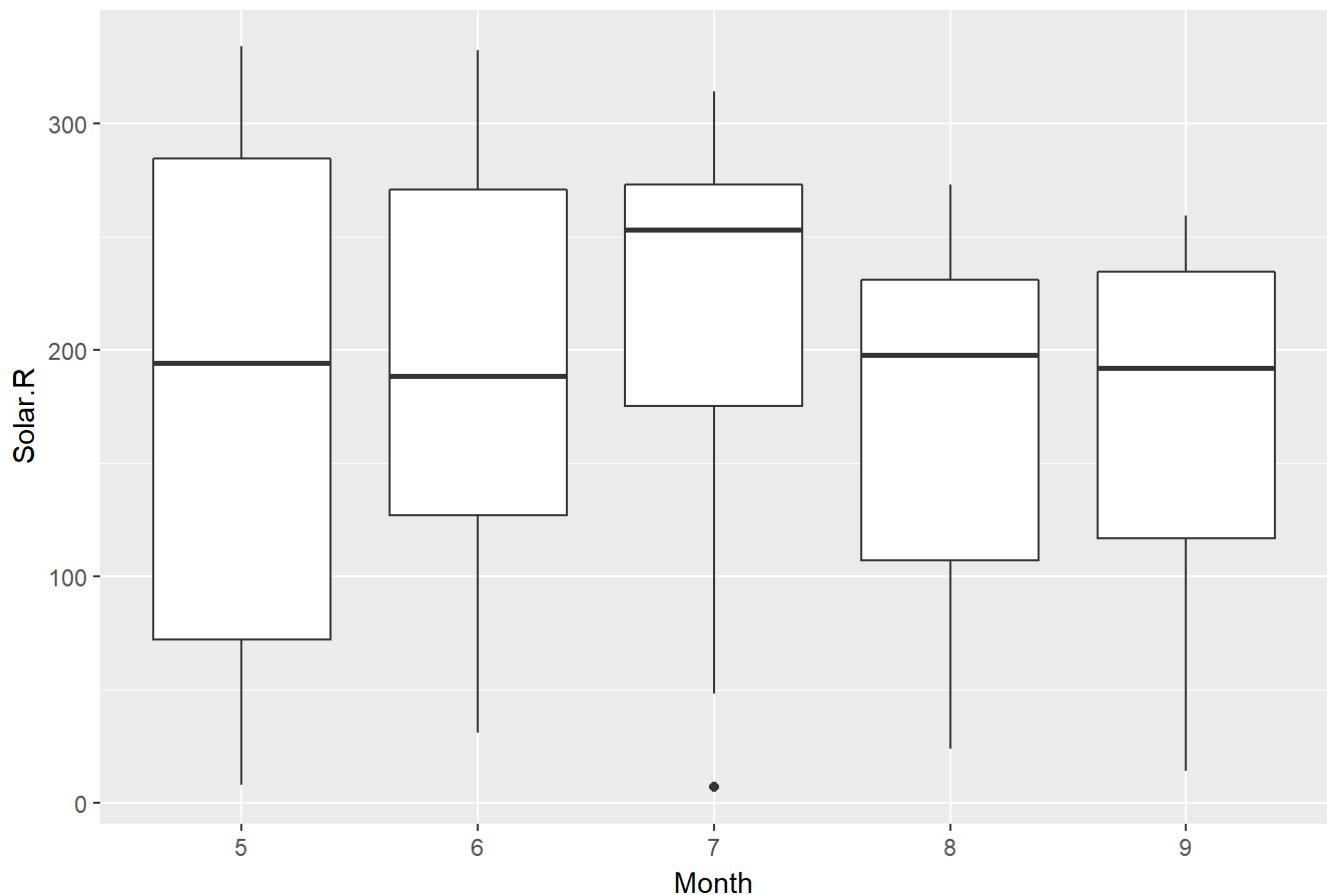
### 3. Boxplot of measurements

```
qplot(Month,Solar.R, data = airquality, binwidth = 2,geom="boxplot",main="Boxplot of Solar.R by Month")
```

```
## Warning: Ignoring unknown parameters: binwidth
```

```
## Warning: Removed 7 rows containing non-finite values (stat_boxplot).
```

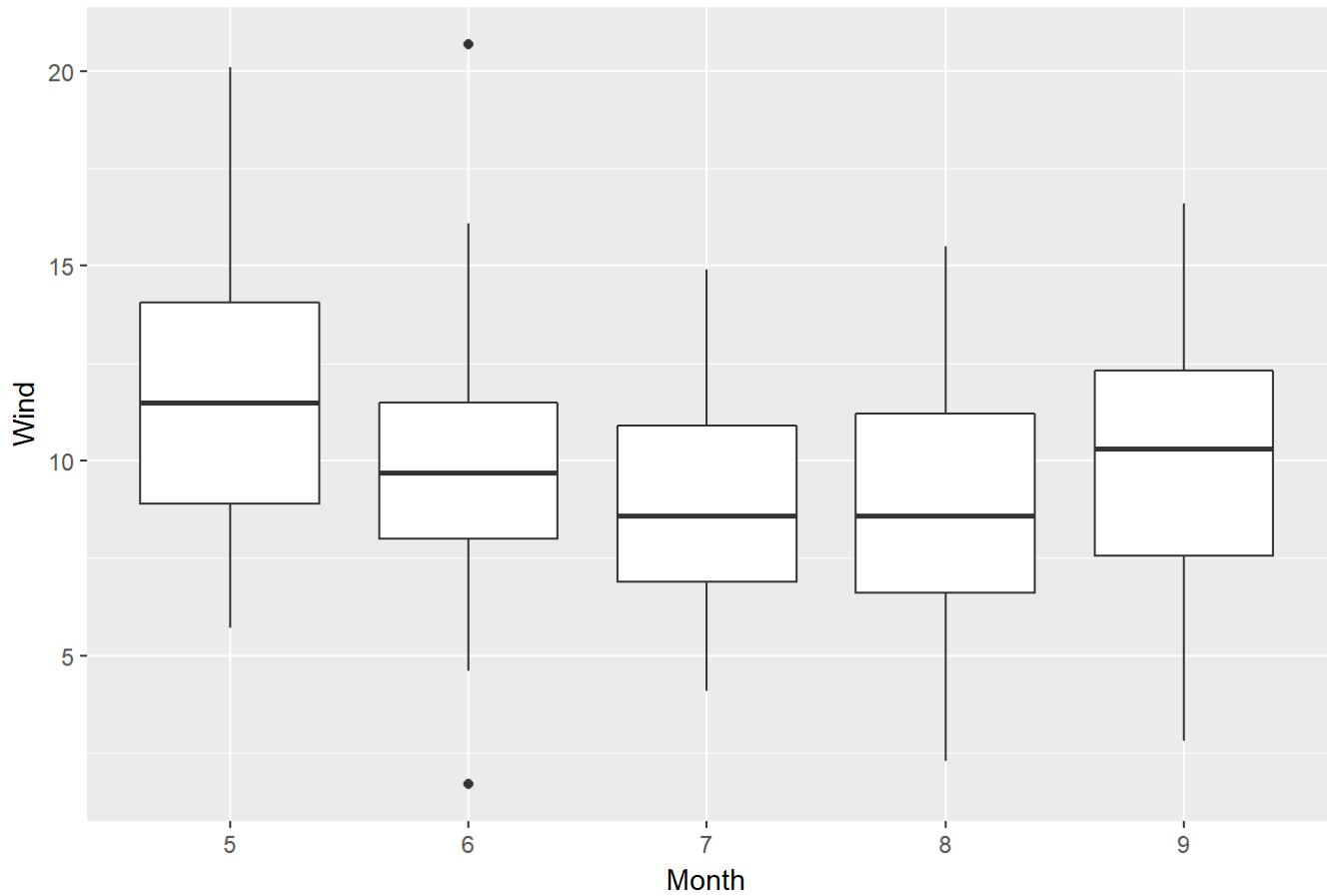
Boxplot of Solar.R by Month



```
qplot(Month,Wind, data = airquality, binwidth = 2,geom="boxplot",main="Boxplot of Wind by Month")
```

```
## Warning: Ignoring unknown parameters: binwidth
```

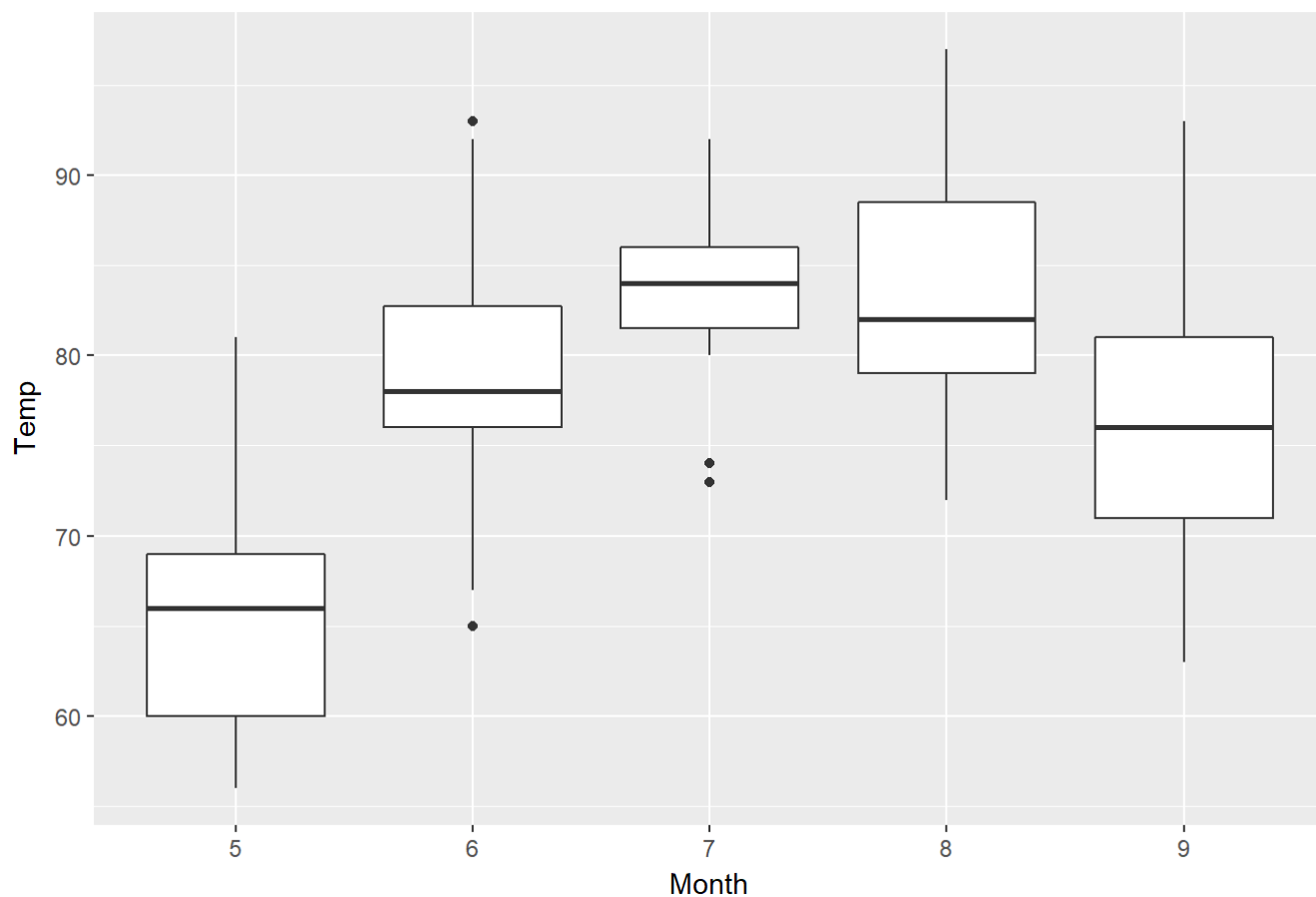
## Boxplot of Wind by Month



```
qplot(Month,Temp, data = airquality, binwidth = 2,geom="boxplot",main="Boxplot of Temperature by Month")
```

```
## Warning: Ignoring unknown parameters: binwidth
```

## Boxplot of Temperature by Month

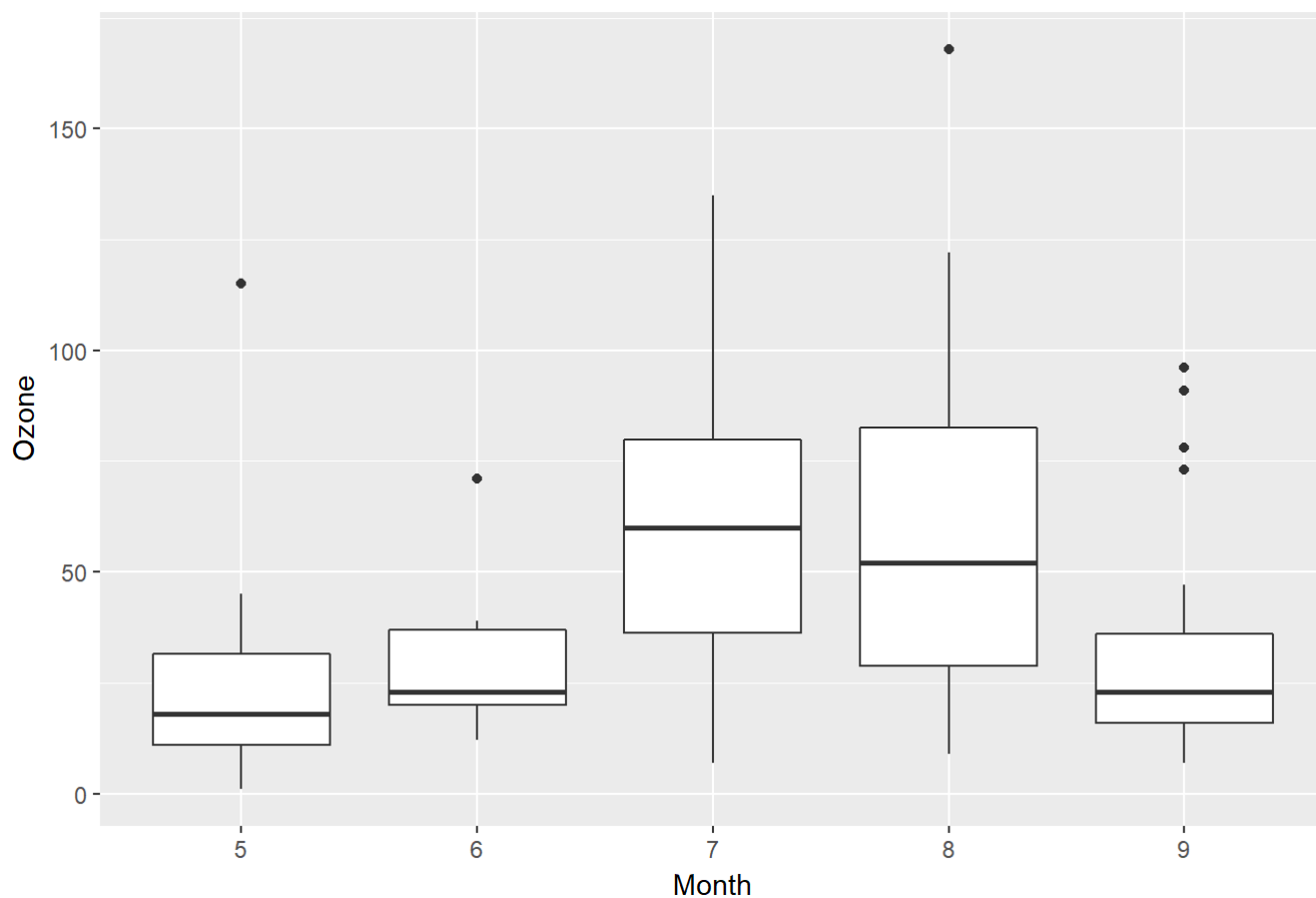


```
qplot(Month,Ozone, data = airquality, binwidth = 2,geom="boxplot",main="Boxplot of Ozone by Month")
```

```
## Warning: Ignoring unknown parameters: binwidth
```

```
## Warning: Removed 37 rows containing non-finite values (stat_boxplot).
```

### Boxplot of Ozone by Month

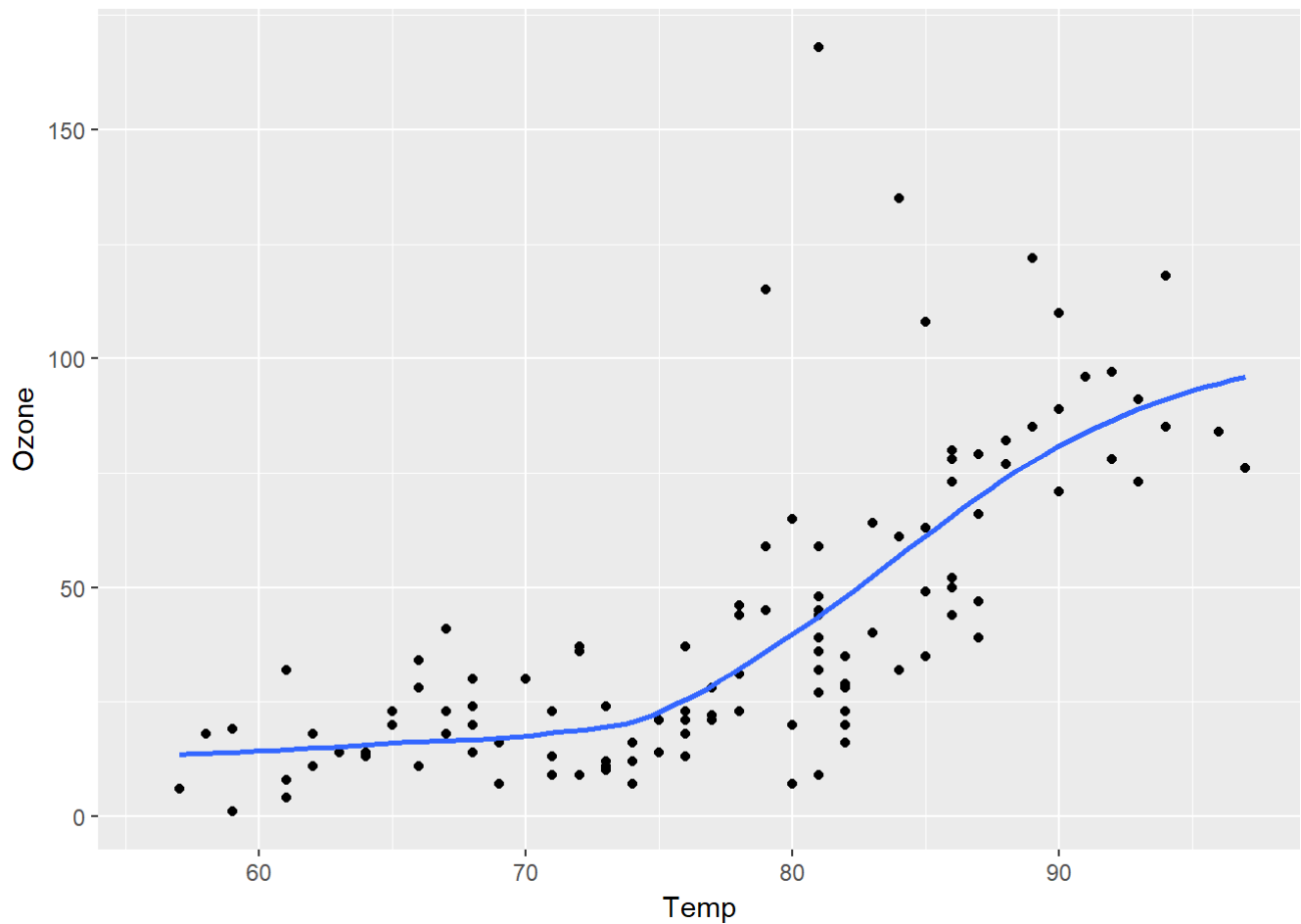


#### 4. ScatterPlot

```
ggplot(airquality, aes(Temp, Ozone)) + geom_point() + geom_smooth(method = "loess", se = FALSE)
```

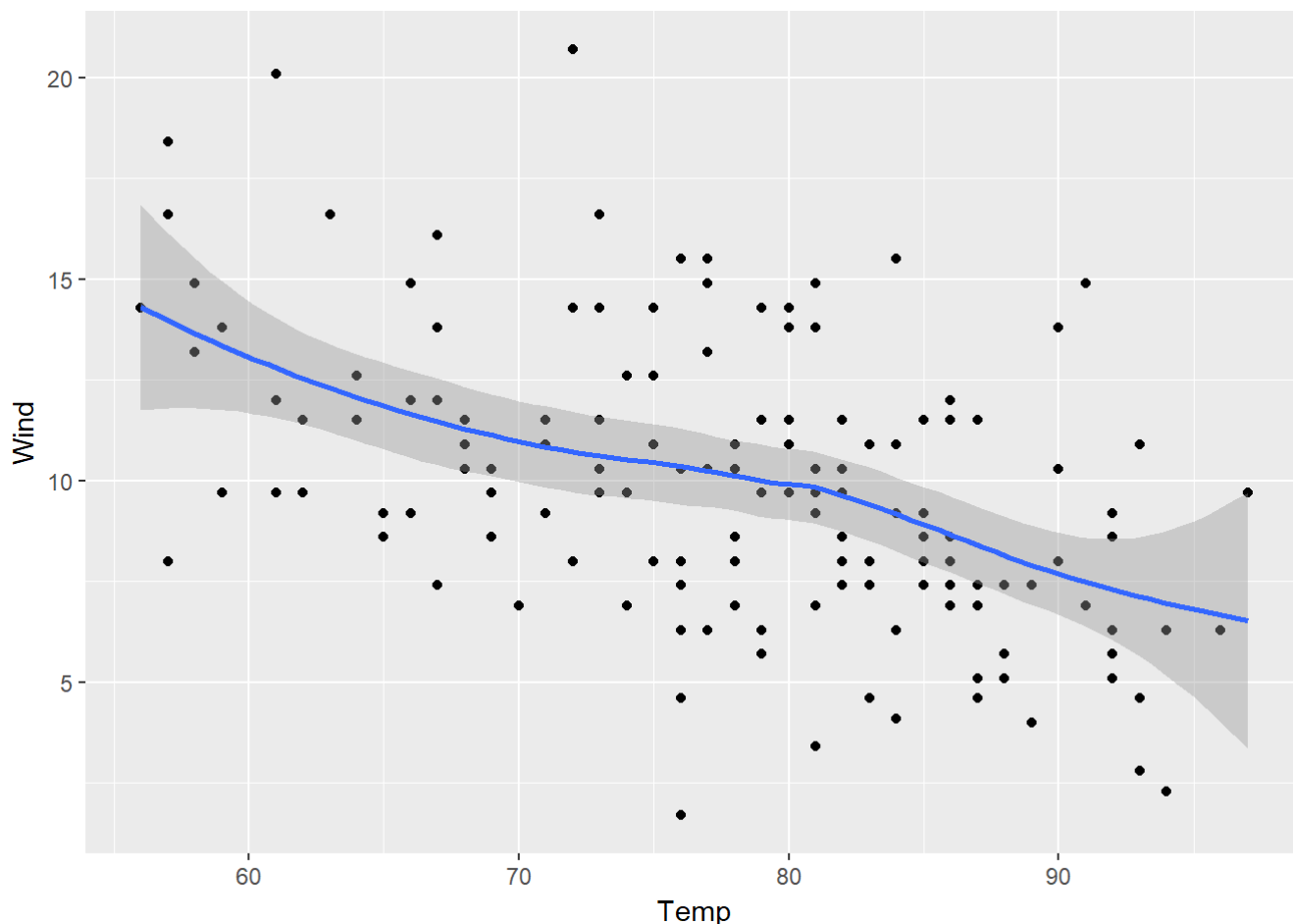
```
## Warning: Removed 37 rows containing non-finite values (stat_smooth).
```

```
## Warning: Removed 37 rows containing missing values (geom_point).
```



```
qplot(Temp, Wind, data = airquality, geom = c("point", "smooth"))
```

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



Summary:

#### Basic plotting

**Advantages:** The base plotting system is often the most convenient plotting system to use because it mirrors how we sometimes think of building plots and analyzing data. It gives you the flexibility to specify these kinds of details to painstaking accuracy. **Downsides:** You would have to set details (e.g. margins) by yourself. Another downside of the base plotting system is that it's difficult to describe or translate a plot to others because there's no clear graphical language or grammar that can be used to communicate what you've done.

#### Lattice system:

**Advantages:** Lattice plots tend to be most useful for conditioning types of plots, i.e. looking at how y changes with x across levels of z. These types of plots are useful for looking at multidimensional data and often allow you to squeeze a lot of information into a single window or page. Another aspect of lattice that makes it different from base plotting is that things like margins and spacing are set automatically. **Downsides:** One downside with the lattice system is that it can sometimes be very awkward to specify an entire plot in a single function call. Also, annotation in panels in plots is not especially intuitive and can be difficult to explain. Finally, once a plot is created, you cannot "add" to the plot,

#### The ggplot2 system:

**Advantages:** The ggplot2 system automatically deals with spacings, text, titles but also allows you to annotate by "adding" to a plot. In a sense, the ggplot2 system takes many of the cues from the base plotting system and formalizes them a bit. Superficially, the ggplot2 functions are similar to lattice, but the system is generally easier and more intuitive to use.