

Bike Shared Data Exploration and Preparation

```
library(ggplot2)
library(dplyr)

## 
## Attaching package: 'dplyr'

## The following objects are masked from 'package:stats':
## 
##     filter, lag

## The following objects are masked from 'package:base':
## 
##     intersect, setdiff, setequal, union

library(magrittr)
library(ggpubr)
library(reshape2)
```

Get Data

```
bike <- read.csv("hour.csv", header= TRUE, )

str(bike)

## 'data.frame':    17379 obs. of  17 variables:
## $ instant    : int  1 2 3 4 5 6 7 8 9 10 ...
## $ dteday     : Factor w/ 731 levels "2011-01-01","2011-01-02",...: 1 1 1 1 1 1 1 1 1 1 ...
## $ season     : int  1 1 1 1 1 1 1 1 1 ...
## $ yr         : int  0 0 0 0 0 0 0 0 0 ...
## $ mnth       : int  1 1 1 1 1 1 1 1 1 ...
## $ hr         : int  0 1 2 3 4 5 6 7 8 9 ...
## $ holiday    : int  0 0 0 0 0 0 0 0 0 ...
## $ weekday    : int  6 6 6 6 6 6 6 6 6 ...
## $ workingday: int  0 0 0 0 0 0 0 0 0 ...
## $ weathersit: int  1 1 1 1 1 2 1 1 1 ...
## $ temp        : num  0.24 0.22 0.22 0.24 0.24 0.24 0.22 0.2 0.24 0.32 ...
## $ atemp       : num  0.288 0.273 0.273 0.288 0.288 ...
## $ hum         : num  0.81 0.8 0.8 0.75 0.75 0.75 0.8 0.86 0.75 0.76 ...
## $ windspeed   : num  0 0 0 0 0.0896 0 0 0 0 ...
## $ casual      : int  3 8 5 3 0 0 2 1 1 8 ...
## $ registered: int  13 32 27 10 1 1 0 2 7 6 ...
## $ cnt         : int  16 40 32 13 1 1 2 3 8 14 ...

head(bike)
```

```

##   instant      dteday season yr mnth hr holiday weekday workingday weathersit
## 1       1 2011-01-01     1 0     1 0      0     6      0      0      1
## 2       2 2011-01-01     1 0     1 1      0     6      0      0      1
## 3       3 2011-01-01     1 0     1 2      0     6      0      0      1
## 4       4 2011-01-01     1 0     1 3      0     6      0      0      1
## 5       5 2011-01-01     1 0     1 4      0     6      0      0      1
## 6       6 2011-01-01     1 0     1 5      0     6      0      0      2
##   temp    atemp   hum windspeed casual registered cnt
## 1 0.24 0.2879 0.81 0.0000     3     13    16
## 2 0.22 0.2727 0.80 0.0000     8     32    40
## 3 0.22 0.2727 0.80 0.0000     5     27    32
## 4 0.24 0.2879 0.75 0.0000     3     10    13
## 5 0.24 0.2879 0.75 0.0000     0      1    1
## 6 0.24 0.2576 0.75 0.0896     0      1    1

```

Clean the Data

```
sum(is.na(bike))
```

```
## [1] 0
```

```
bike$instant<- NULL
```

```
library(weathermetrics)
```

```
bike <- mutate(bike, temp = fahrenheit.to.celsius(bike$temp *100 ))
```

```
bike <- mutate(bike, atemp = fahrenheit.to.celsius(bike$atemp *100 ))
```

convert temperatur from F/100 to C for better caculation

convert date from numeral feature to datetime feature

```
bike$dteday <- as.POSIXct(bike$dteday)
```

Exploratory Data Analysis

```
head(bike)
```

```

##      dteday season yr mnth hr holiday weekday workingday weathersit temp
## 1 2011-01-01     1 0    1 0     0     6      0      1 -4.44
## 2 2011-01-01     1 0    1 1     0     6      0      1 -5.56
## 3 2011-01-01     1 0    1 2     0     6      0      1 -5.56
## 4 2011-01-01     1 0    1 3     0     6      0      1 -4.44
## 5 2011-01-01     1 0    1 4     0     6      0      1 -4.44
## 6 2011-01-01     1 0    1 5     0     6      0      2 -4.44
##   atemp  hum windspeed casual registered cnt
## 1 -1.78 0.81  0.0000   3    13  16
## 2 -2.63 0.80  0.0000   8    32  40
## 3 -2.63 0.80  0.0000   5    27  32
## 4 -1.78 0.75  0.0000   3    10  13
## 5 -1.78 0.75  0.0000   0     1  1
## 6 -3.47 0.75  0.0896   0     1  1

```

```
colnames(bike)
```

```

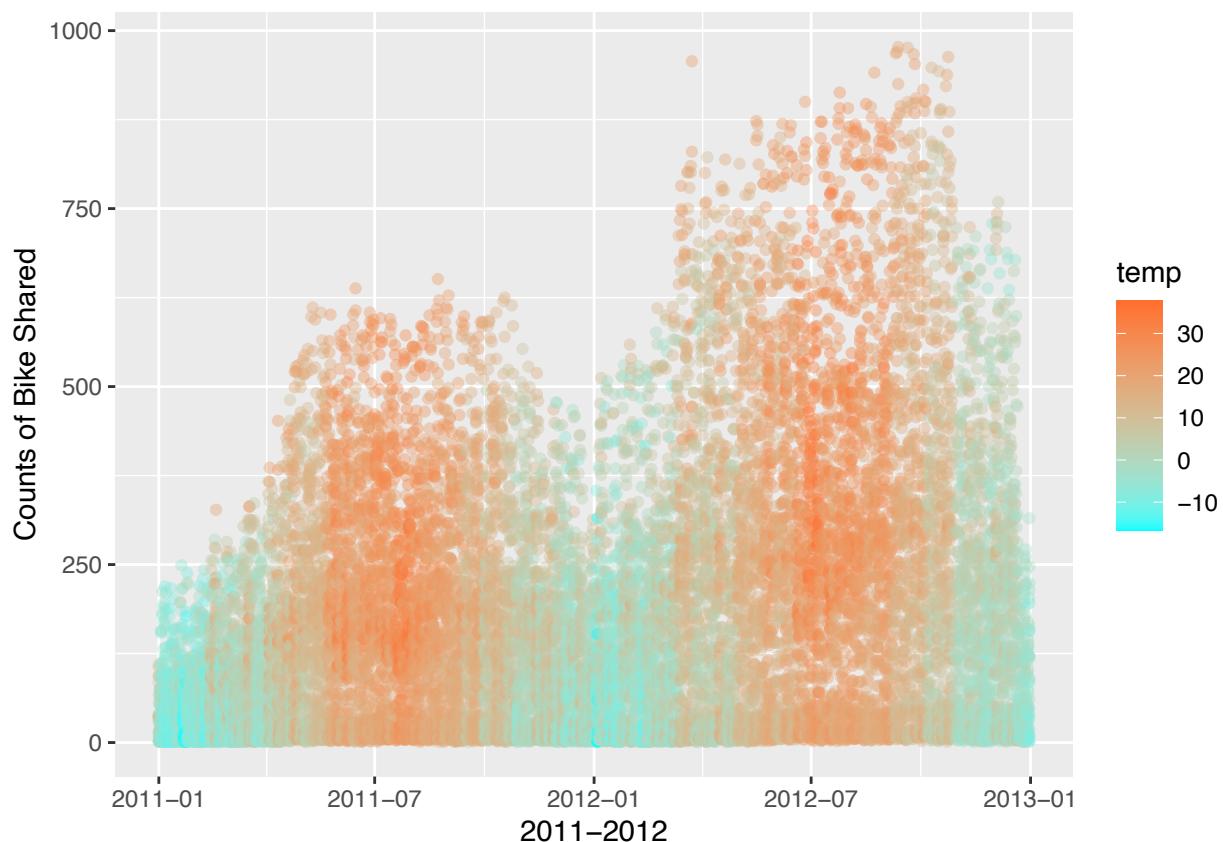
## [1] "dteday"      "season"       "yr"          "mnth"        "hr"
## [6] "holiday"     "weekday"     "workingday"   "weathersit"   "temp"
## [11] "atemp"       "hum"         "windspeed"    "casual"      "registered"
## [16] "cnt"

```

```

ggplot(bike,aes(dteday,cnt)) +
  geom_point(aes(color=temp),alpha=0.4) +
  scale_color_continuous(low="#85d2d4", high = "#FF6E2E") +
  xlab("2011-2012") + ylab("Counts of Bike Shared")

```



```

which(bike$dteday == "2012-01-01")

## [1] 8646 8647 8648 8649 8650 8651 8652 8653 8654 8655 8656 8657 8658 8659 8660
## [16] 8661 8662 8663 8664 8665 8666 8667 8668 8669

```

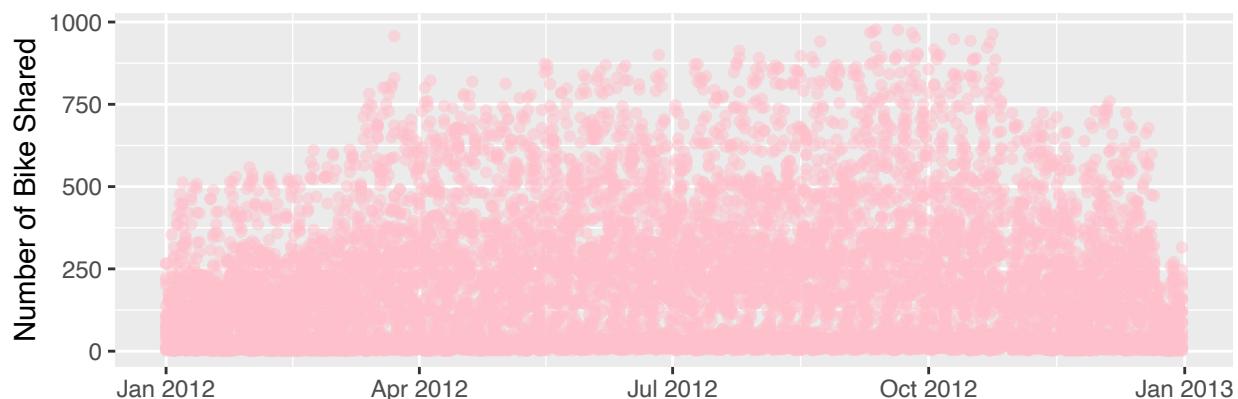
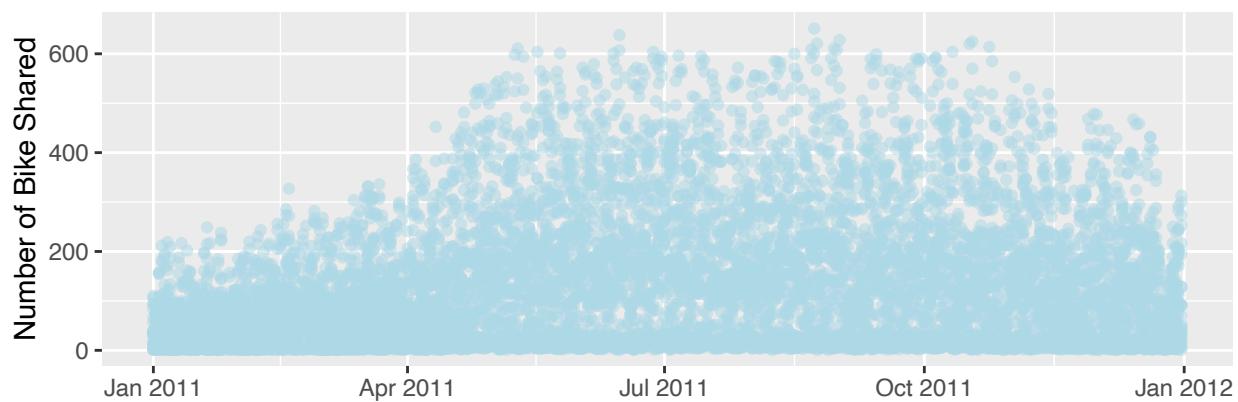
```
bike2011 <- bike[1:8645, ]
```

```
bike2012<- bike[8646:17379, ]
```

```
b2011<- ggplot(bike2011, aes(x= dteday, y = cnt)) +
  geom_point(color="light blue",alpha= 0.5) +
  xlab(element_blank()) + ylab("Number of Bike Shared")
```

```
b2012<- ggplot(bike2012, aes(x= dteday, y = cnt)) +
  geom_point(color="pink",alpha= 0.5) +
  xlab(element_blank()) + ylab("Number of Bike Shared")
```

```
ggarrange(b2011,b2012 ,
          ncol = 1, nrow = 2)
```



```
bikecasual<- ggplot(bike2011, aes(x= dteday, y = casual)) +
  geom_point(color="light green",alpha= 0.4) +
  xlab(element_blank())+ ylab("Number of Bike Shared") +
```

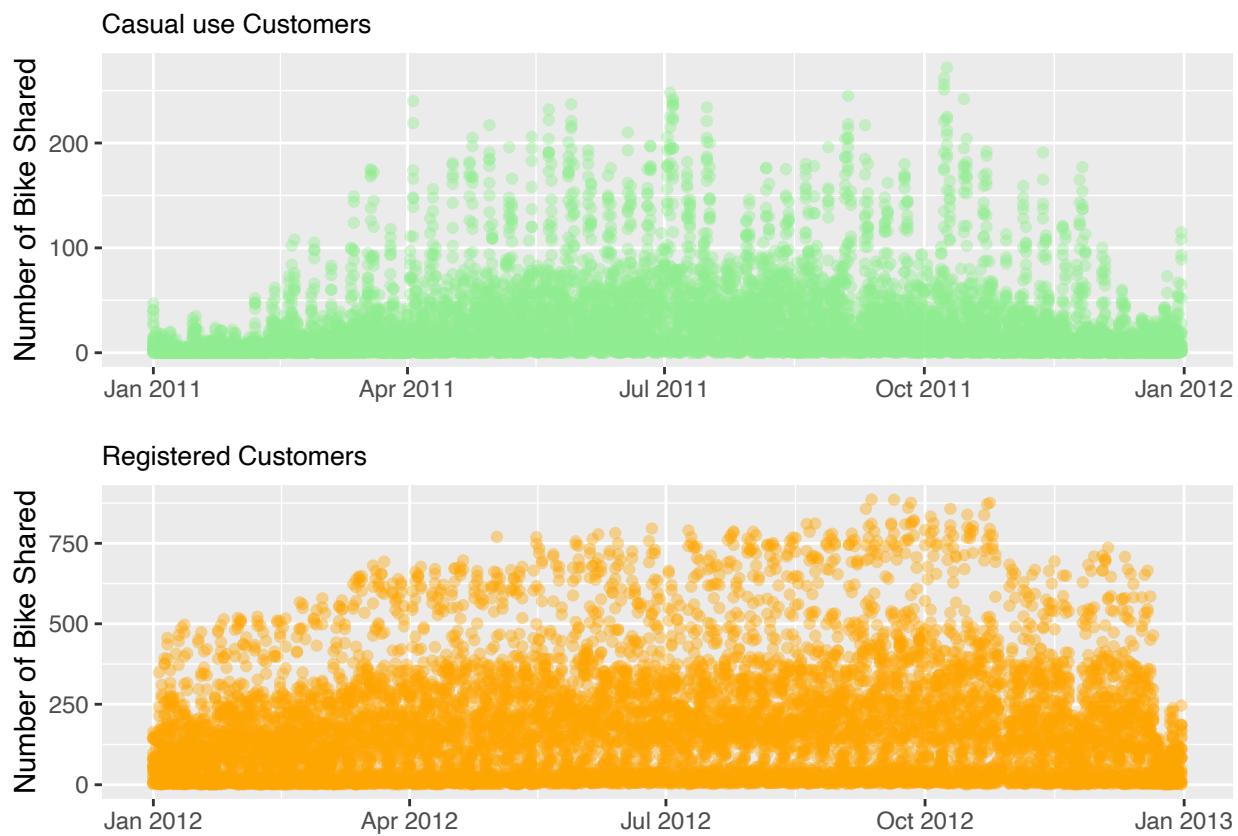
```

ggtitle("Casual use Customers") +
  theme(plot.title = element_text(size = 10))

bikeregistered <- ggplot(bike2012, aes(x= dteday, y = registered)) +
  geom_point(color="orange",alpha= 0.4) +
  xlab(element_blank()) + ylab("Number of Bike Shared") +
  ggtitle("Registered Customers") +
  theme(plot.title = element_text(size = 10))

ggarrange(bikecasual,bikeregistered,
  ncol = 1, nrow = 2)

```

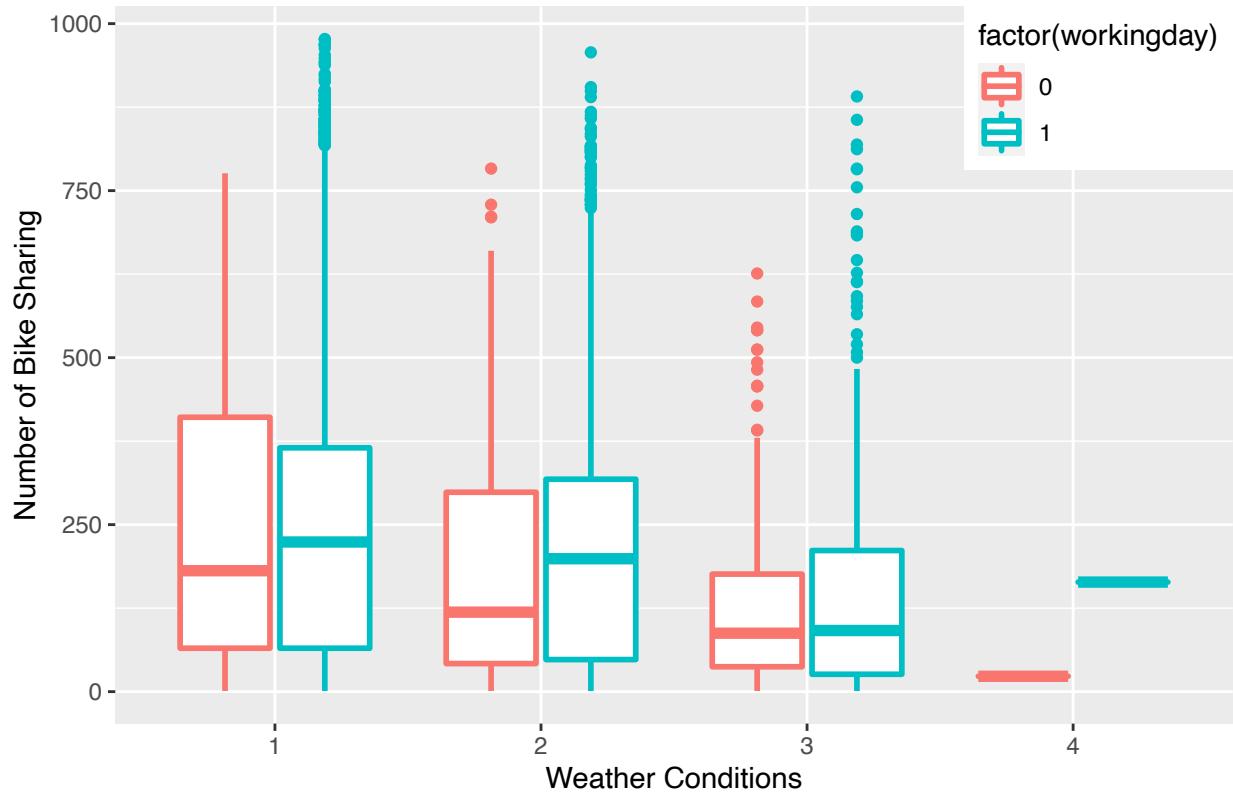


```

ggplot(bike2012, aes(x= factor(weather), y = cnt)) +
  geom_boxplot(aes(color=factor(workingday)), size =1) +
  xlab("Weather Conditions") + ylab("Number of Bike Sharing") +
  ggtitle(" Bike Sharing vs Weather Situation Base on Workingday") +
  theme( legend.position = c(1,1), legend.justification = c(1,1))

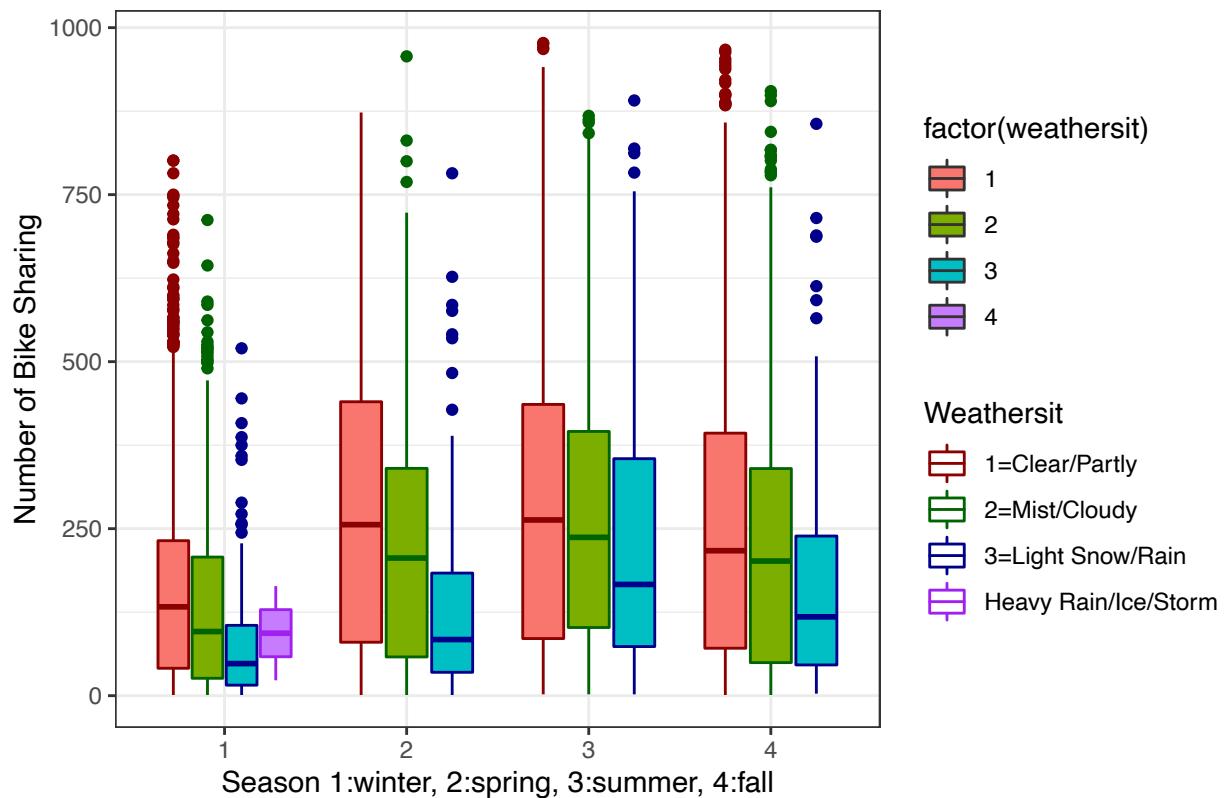
```

Bike Sharing vs Weather Situation Base on Workingday

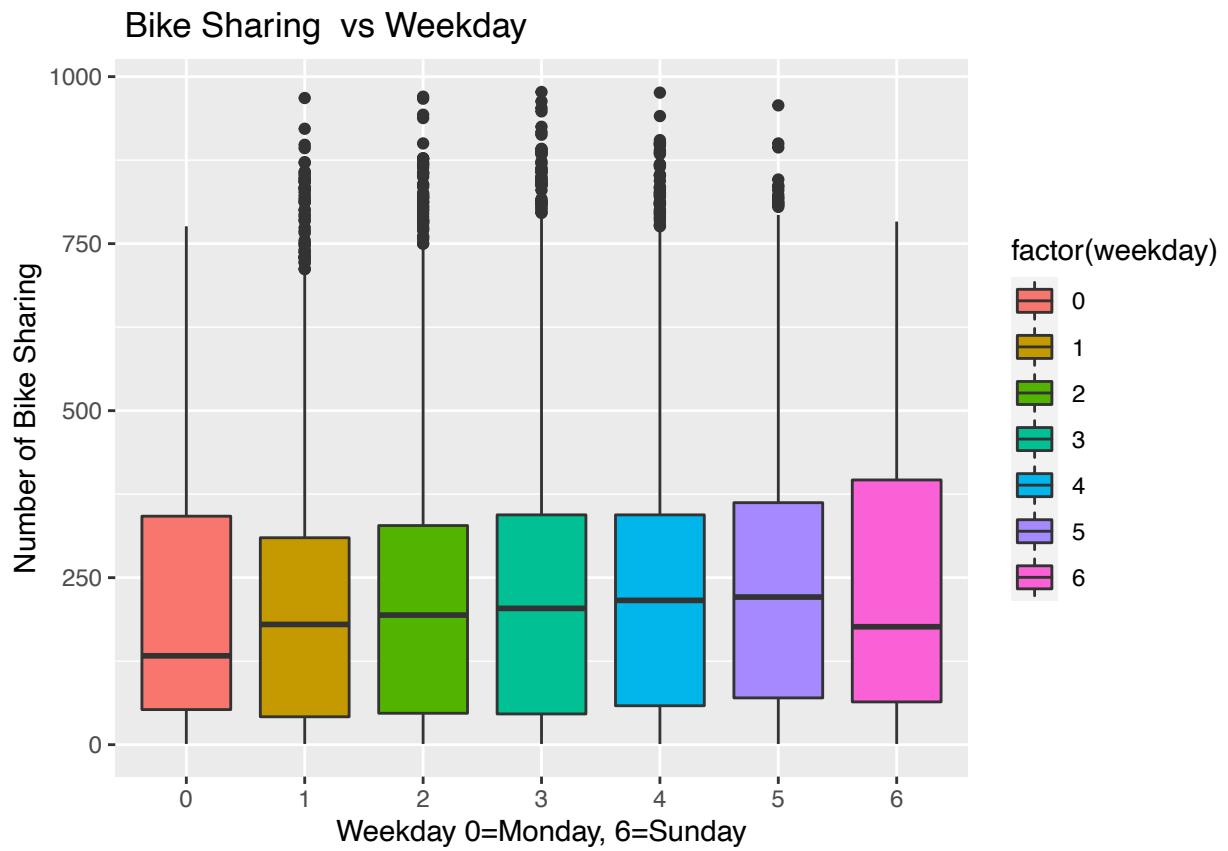


```
ggplot(bike2012, aes(x=factor(season), y = cnt, color= factor(weather)))+
  geom_boxplot(aes(fill=factor(weather))) +
  labs(title =" Bike Sharing vs Weather Condition in each Season", x = "Season 1:winter, 2:spring, 3:summer, 4:fall", y = "Bike Sharing")+
  scale_color_manual(labels = c("1=Clear/Partly", "2=Mist/Cloudy", "3=Light Snow/Rain","Heavy Rain/Ice/Wind"), values = c( "dark red", "dark green","dark blue", "purple"))+
  theme_bw()+
  guides(color=guide_legend("Weather"))
```

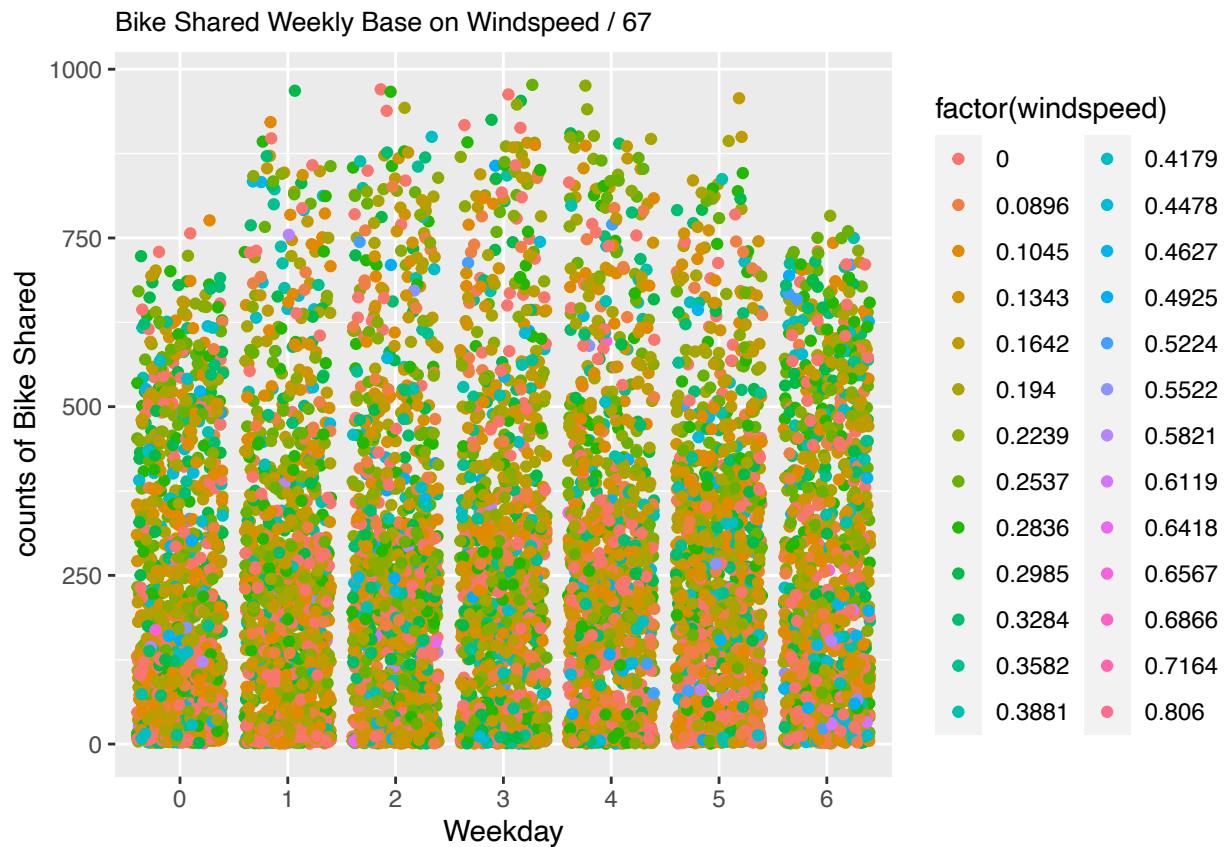
Bike Sharing vs Weather Condition in each Season



```
ggplot(bike2012, aes(x=factor(weekday), y = cnt)) +
  geom_boxplot(aes(fill = factor(weekday))) +
  xlab("Weekday 0=Monday, 6=Sunday") + ylab("Number of Bike Sharing") +
  ggtitle(" Bike Sharing vs Weekday")
```

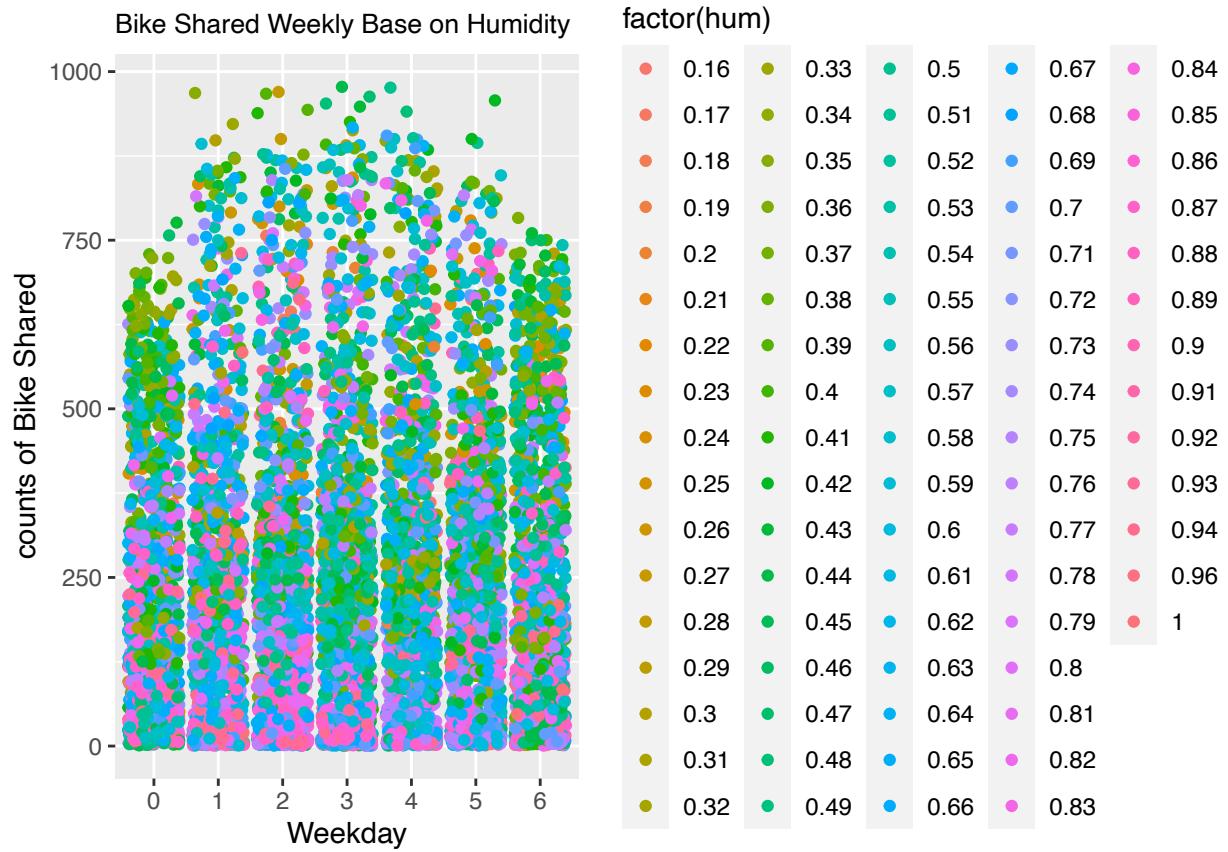


```
ggplot(bike2012, aes(x = factor(weekday), y = cnt, colour = factor(windspeed))) +
  geom_point(position = position_jitter()) +
  xlab("Weekday") + ylab("counts of Bike Shared") +
  ggtitle("Bike Shared Weekly Base on Windspeed / 67") +
  theme(plot.title = element_text(size = 10))
```



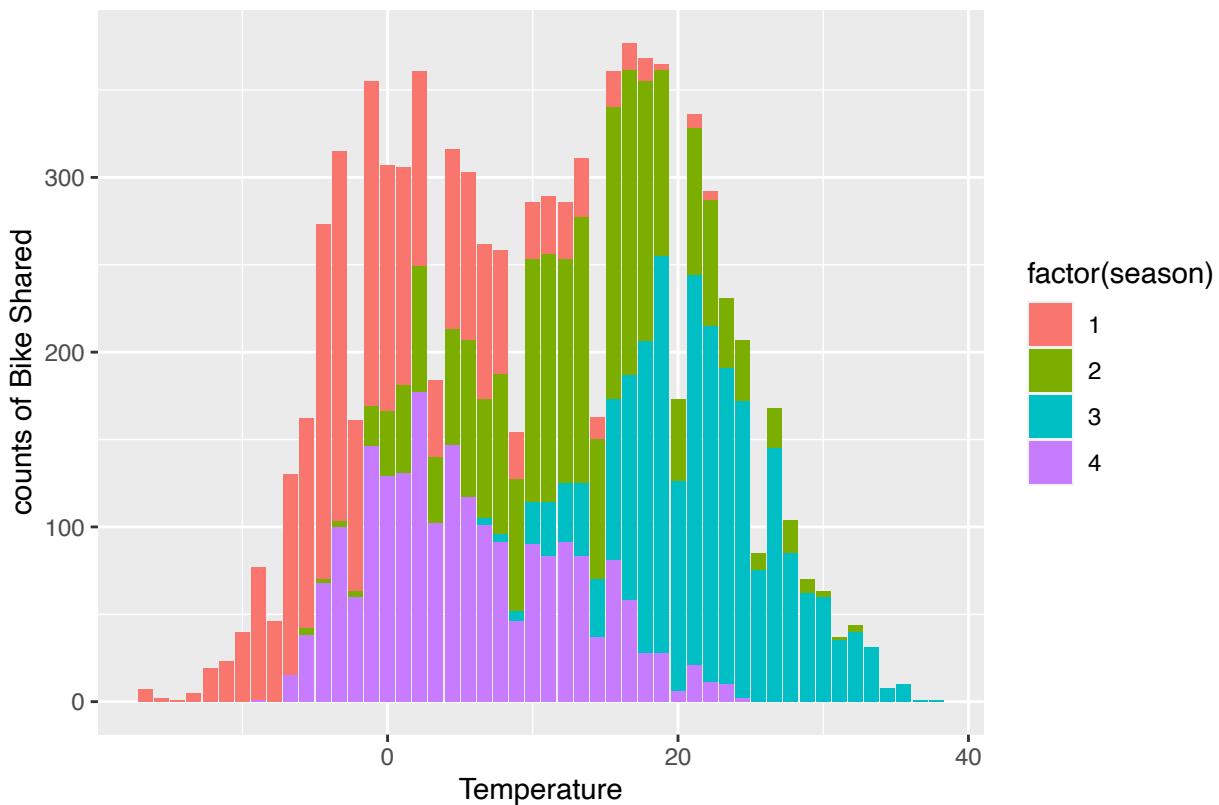
The windspeed in the raw dataset have been devided by 67.

```
ggplot(bike2012, aes(x = factor(weekday), y = cnt, colour = factor(hum))) +
  geom_point(position = position_jitter()) +
  xlab("Weekday") + ylab("counts of Bike Shared") +
  ggtitle("Bike Shared Weekly Base on Humidity") +
  theme(plot.title = element_text(size = 10))
```



```
ggplot(bike2012, aes(x = temp)) +
  geom_bar(aes(fill=factor(season))) +
  xlab("Temperature") + ylab("counts of Bike Shared") +
  ggtitle("Bike Shared vs Temperature each Season") +
  theme(plot.title = element_text(size = 10))
```

Bike Shared vs Temperature each Season



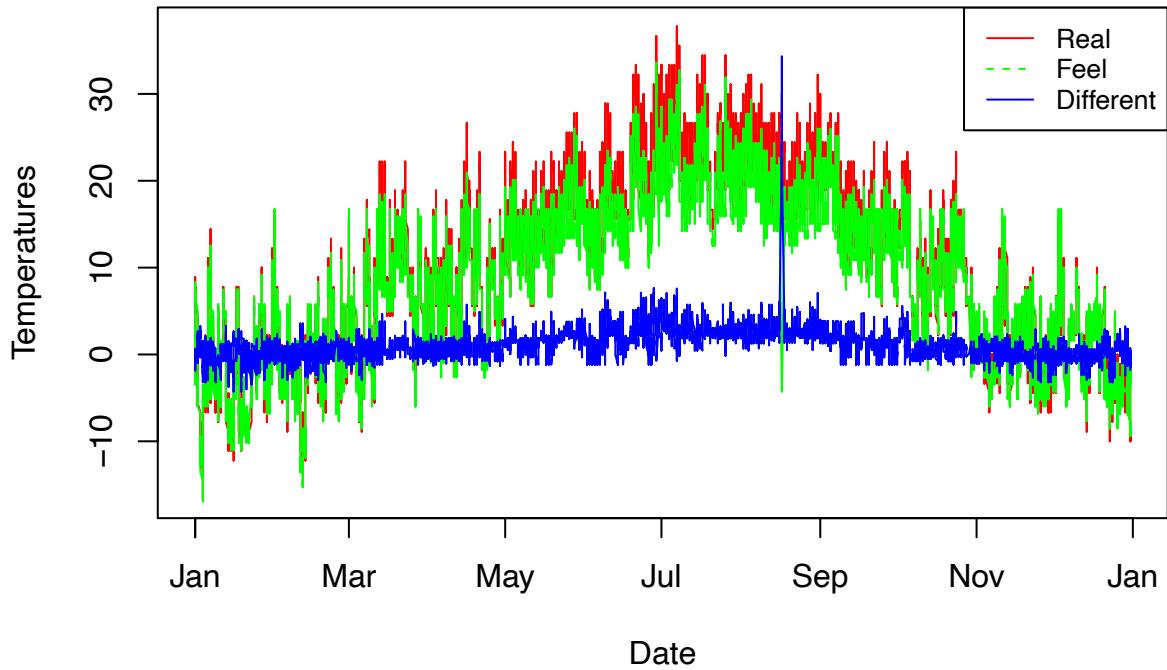
```
bike2012$dtemp <- (bike2012$temp - bike2012$atemp)
```

```
mean(bike2012$dtemp)
```

```
## [1] 1.24178
```

```
plot(bike2012$dteday, bike2012$temp, type="l", col="red", xlab="Date", ylab="Temperatures", main= "Real, Feel, Different")
lines(bike2012$dteday, bike2012$atemp, type="l", col="green")
lines(bike2012$dteday, bike2012$dtemp, type="l", col="blue")
legend("topright", legend=c("Real", "Feel", "Different"), col=c("red", "green", "blue"), lty = 1:2, cex=0.8)
```

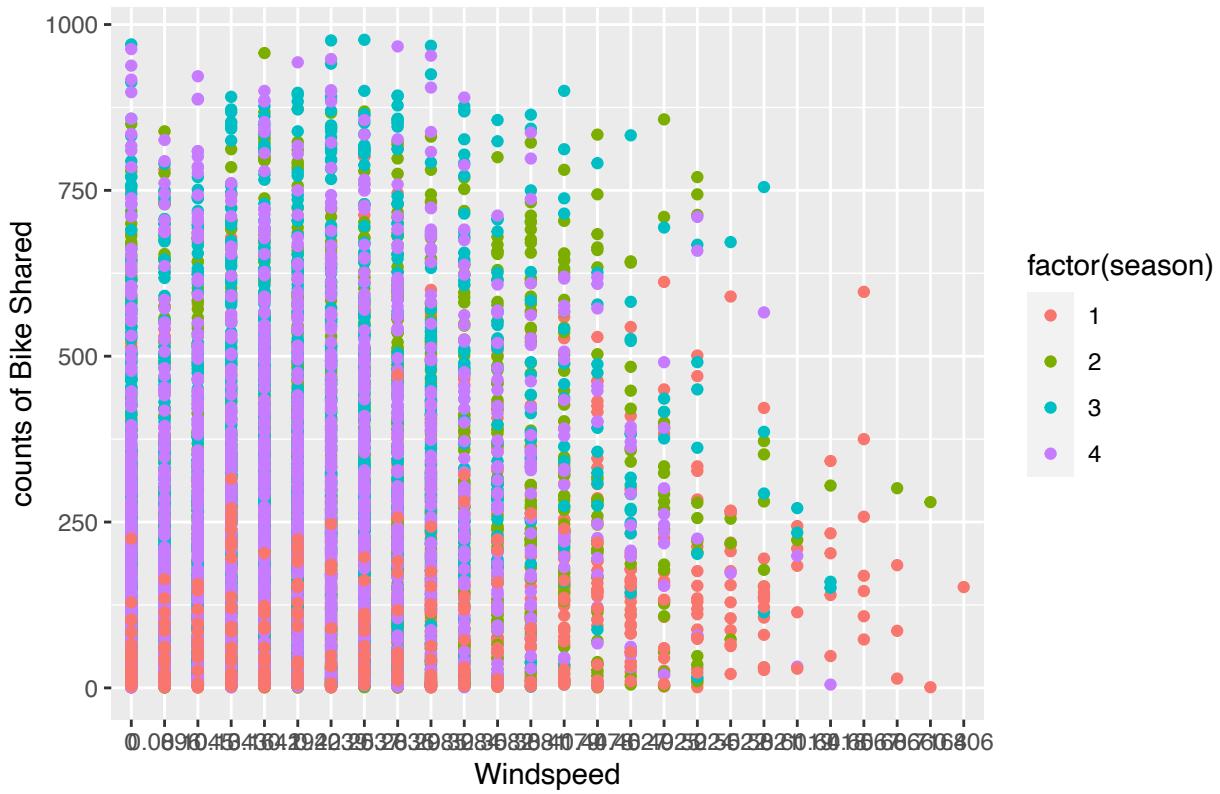
Real, Feel and Different Temperatures



```
bike2012$dtemp <- NULL
```

```
ggplot(bike2012, aes(x=factor(windspeed), y= cnt)) +  
  geom_point(aes(color=factor(season))) +  
  xlab("Windspeed") + ylab("counts of Bike Shared") +  
  ggtitle("Bike Shared vs Windspeed Base on Season") +  
  theme(plot.title = element_text(size = 15))
```

Bike Shared vs Windspeed Base on Season



```

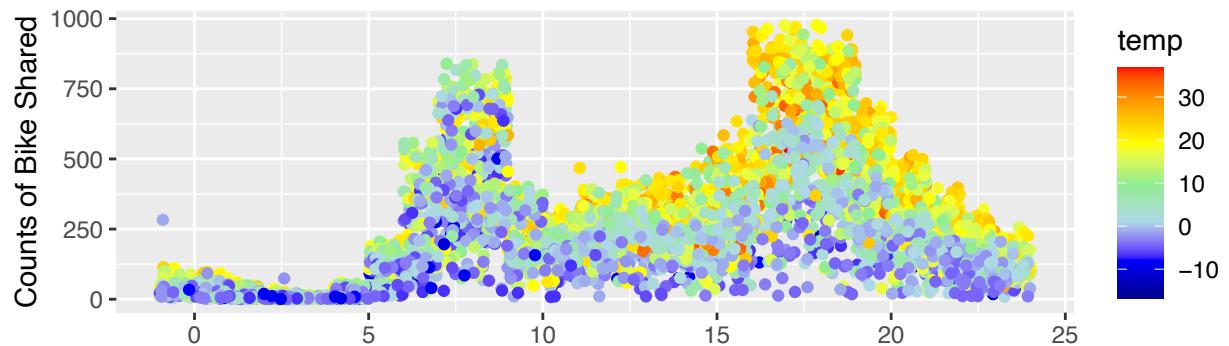
weekd <- ggplot(filter(bike2012, workingday== 1), aes(hr,cnt)) +
  geom_point(position = position_jitter(w=1,h=0),aes(color=temp)) +
  scale_colour_gradientn(colours = c("dark blue","blue", "light blue", "light green", "yellow","orange"))
  xlab(element_blank())+ ylab("Counts of Bike Shared") +
  ggtitle("Bike Shared vs Hours Base on Temperature in Workingday") +
  theme(plot.title = element_text(size = 15))

wend<- ggplot(filter(bike2012, workingday== 0), aes(hr,cnt)) +
  geom_point(position = position_jitter(w=1,h=0),aes(color=temp)) +
  scale_colour_gradientn(colours = c("dark blue","blue", "light blue", "light green", "yellow","orange"))
  xlab("Hours (0-24)")+ ylab("Counts of Bike Shared") +
  ggtitle("Bike Shared vs Hours Base on Temperature in None Workingday") +
  theme(plot.title = element_text(size = 14))

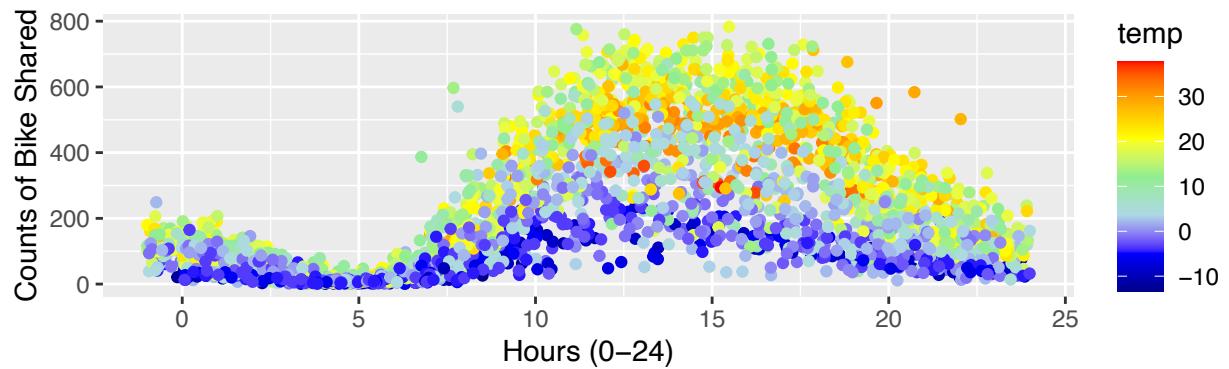
ggarrange(weekd, wend,
          ncol = 1, nrow = 2)

```

Bike Shared vs Hours Base on Temperature in Workingday

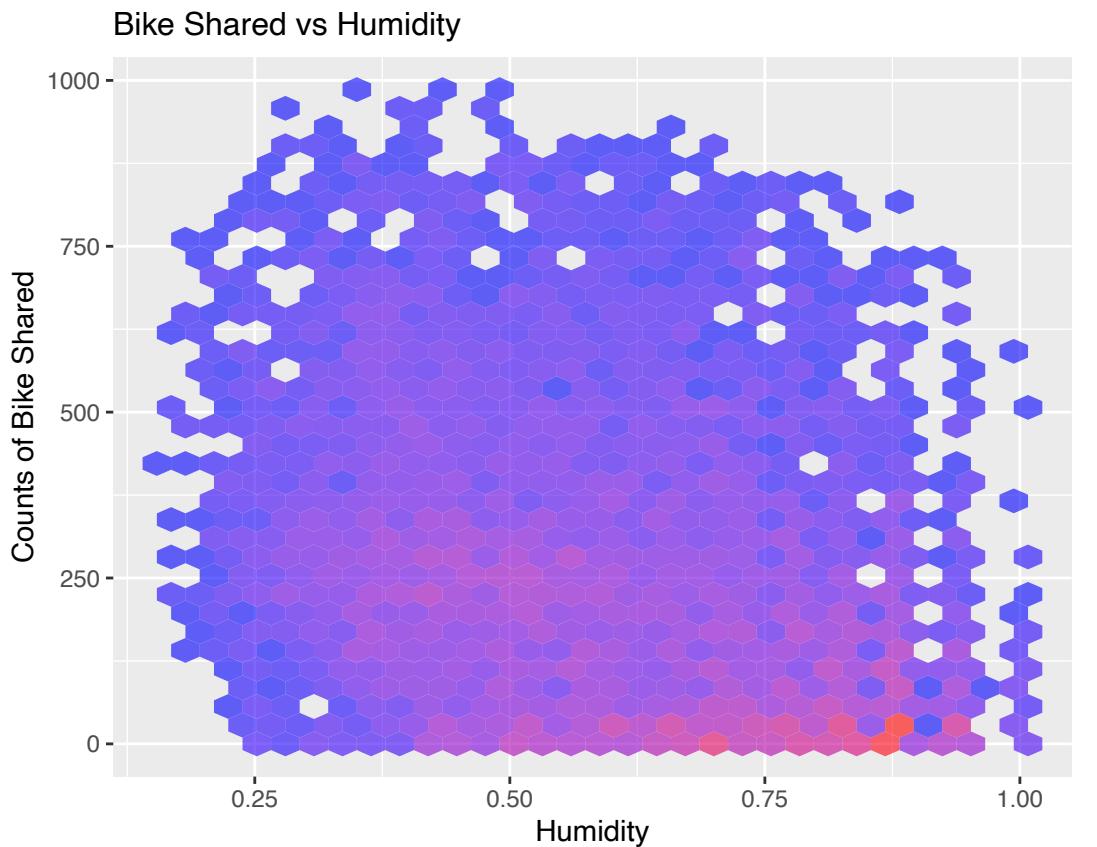


Bike Shared vs Hours Base on Temperature in None Workingday



```
library(hexbin)
```

```
ggplot(bike2012, aes(x=hum, y = cnt)) +  
  geom_hex(alpha=0.6) + scale_fill_gradient(high = "red", low = "blue") +  
  xlab("Humidity") + ylab("Counts of Bike Shared") +  
  ggtitle("Bike Shared vs Humidity") +  
  theme(plot.title = element_text(size = 12))
```



```
pl <- ggplot(bike2012, aes(x= temp, y = cnt)) +
  geom_point(aes(shape=factor(season),color=factor(season)), size =1) +
  xlab("Temprature") + ylab("Number of Bike Sharing") +
  ggtitle(" Bike Sharing vs Temperatures")
print(pl)
```

Bike Sharing vs Temperatures

