Bike sharing system research

Gurgen Blbulyan

10/17/2019

First of all let's see on which data I do research. I take a Capital Bikeshare company data ,which is bicycle sharing system in Washington D.C..

To see what is contained data, let's do some tricks.

```
bike<-read.csv("bikeshare.csv")
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 1 ...
##  $ yr        : int  0 0 0 0 0 0 0 0 0 0 ...
##  $ mnth      : int  1 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 0 ...
##  $ weekday   : int  6 6 6 6 6 6 6 6 6 6 ...
##  $ workingday: int  0 0 0 0 0 0 0 0 0 0 ...
##  $ weathersit: int  1 1 1 1 1 2 1 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 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 ...
```

As we see we have 17379 observations and 17 variables. Let's see what does our variables showe.

instant: record index , dteday : date , season : season (1:winter, 2:spring, 3:summer, 4:fall) , yr : year (0: 2011, 1:2012) , mnth : month (1 to 12) , hr : hour (0 to 23) , holiday : weather day is holiday or not , weekday : day of the week , workingday : if day is neither weekend nor holiday is 1, otherwise is 0. , weathersit : -1: Clear, Few clouds, Partly cloudy, Partly cloudy -2: Mist + Cloudy, Mist + Broken clouds, Mist + Few clouds, Mist -3: Light Snow, Light Rain + Thunderstorm + Scattered clouds, Light Rain + Scattered clouds -4: Heavy Rain + Ice Pallets + Thunderstorm + Mist, Snow + Fog , temp : Normalized temperature in Celsius. The values are derived via (t-t_min)/(t_max-t_min), t_min=-8, t_max=+39, atemp: Normalized feeling temperature in Celsius. The values are derived via (t-t_min)/(t_max-t_min), t_min=-16, t_max=+50, hum: Normalized humidity. The values are divided to 100 (max) , windspeed: Normalized wind speed. The values are divided to 67 (max) , casual: count of casual users , registered: count of registered users , cnt: count of total rental bikes including both casual and registered

To make our data more compact let's drope some variables.

```
bike$yr=NULL
bike$dteday=NULL
bike$holiday=NULL
bike$temp=NULL
bike$instant=NULL
bike$casual=NULL
```

```
bike$registered=NULL
str(bike)
```

```
## 'data.frame':    17379 obs. of  10 variables:
##  $ season    : int  1 1 1 1 1 1 1 1 1 1 ...
##  $ mnth      : int  1 1 1 1 1 1 1 1 1 1 ...
##  $ hr        : int  0 1 2 3 4 5 6 7 8 9 ...
##  $ weekday   : int  6 6 6 6 6 6 6 6 6 6 ...
##  $ workingday: int  0 0 0 0 0 0 0 0 0 0 ...
##  $ weathersit: int  1 1 1 1 1 2 1 1 1 1 ...
##  $ 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 0.0896 0 0 0 0 ...
##  $ cnt       : int  16 40 32 13 1 1 2 3 8 14 ...
```

Now we have 10 variables now.

Let's import some essential libraries.

```
options(repos="https://cran.rstudio.com" )
library(ggplot2)
library(devtools)
```

```
Loading required package: usethis
```

```
install_github("Avetik666/MartirosSaryan")
```

```
Skipping install of 'martirossaryan' from a github remote, the SHA1 (076f8ec5) has not changed since las
  Use `force = TRUE` to force installation
```

```
install.packages("readbitmap")
```

```
The downloaded binary packages are in
    /var/folders/g_/lggcf4d57rd12w9yvck7l5zh0000gn/T//Rtmpjep1mq/downloaded_packages
```

```
library(readbitmap)
library(martirossaryan)
```

To color our plots in Saryans paintings colors, I create a theme

```
getPalleteNames()
```

```
##  [1] "a-street-constantinople" "a-street-evening"
##  [3] "a-street-noon"           "apple-trees-in-blossom"
##  [5] "armenia"                 "armenian-landscape"
##  [7] "at-the-well"             "blue-pitcher"
##  [9] "by-the-sea"              "constantinople-1"
## [11] "constantinople"          "desert"
## [13] "dogs"                    "egyptian-masks"
## [15] "egyptian-night"          "egyptian-women"
## [17] "fellah-village"          "flowers-from-chamlych"
## [19] "flowers-of-kalaki"       "flowers"
## [21] "fruit-shop"              "grapes"
## [23] "green-jug-and-bouquet"   "heat-running-dog"
## [25] "holiday-under-the-tree"  "hyenas"
## [27] "in-the-grove-in-sambek"  "in-the-shadow"
## [29] "lotus"                   "love"
```

```
## [31] "morning-in-stavrin"      "night-landscape-egypt"
## [33] "persian-woman"           "portrait-of-i-shchukin"
## [35] "red-flowers"             "self-portrait-1"
## [37] "self-portrait"           "steppe-flowers"
## [39] "still-life-bananas"      "still-life-fruits"
## [41] "street-in-cairo"         "to-the-spring"
## [43] "trees-in-blossom"        "violet-jug"
## [45] "wildflowers"             "wisteria"
```

```
getSaryanPallete("armenian-landscape")
```

```
##  [1] "#AA8C3C" "#FBB965" "#214E5F" "#B7D6AE" "#7B714B" "#98840A" "#FBC583"
##  [8] "#316E93" "#B2B3B9" "#FDDC3D" "#BAA692" "#FBB003" "#804F15" "#F4B035"
## [15] "#E89507" "#A5C7F3" "#4B3418" "#0C1514" "#E5E4E1" "#F3C9A4" "#FDE470"
## [22] "#C69311" "#77889B" "#BA9768" "#D8B279" "#FCE2A1" "#8C8D77" "#FCC604"
## [29] "#919CAD" "#D59D41" "#576B2C" "#8AB1E1"
```

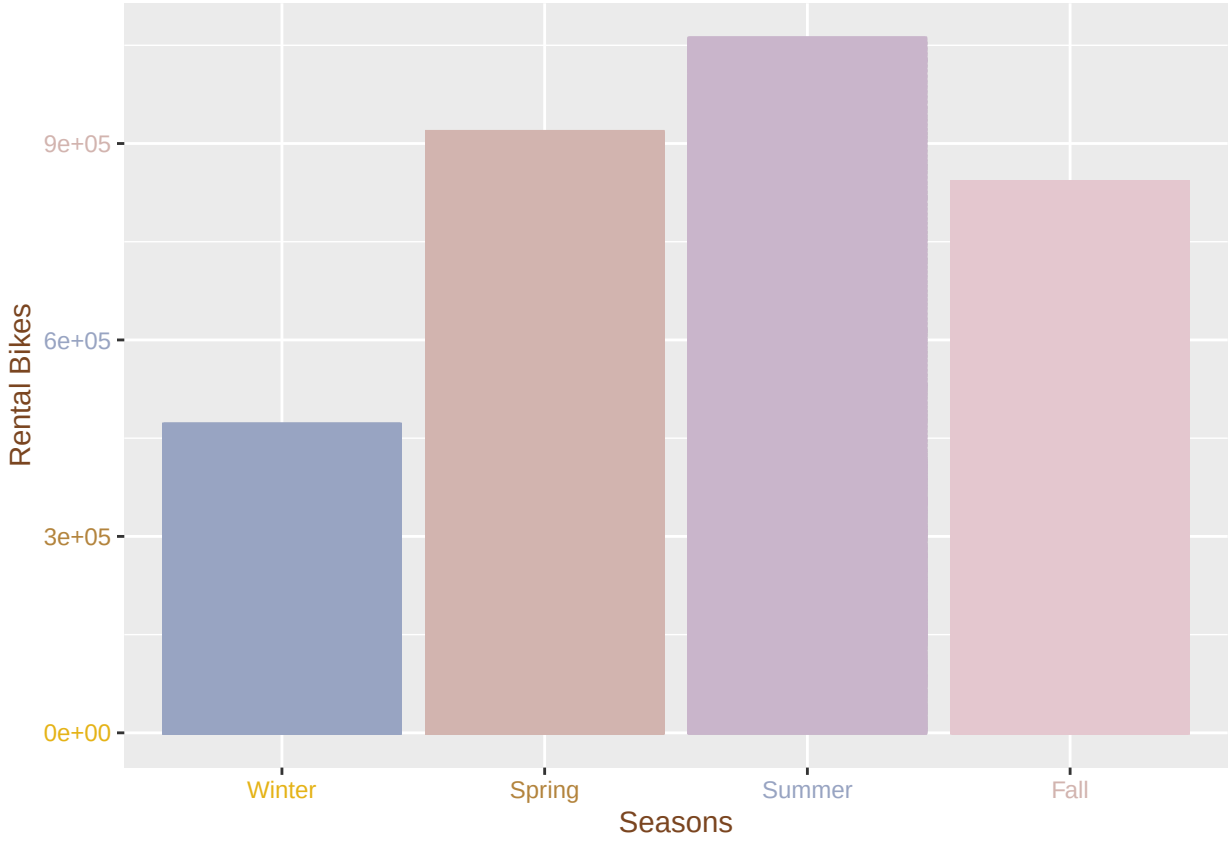
```
getSaryanPallete("armenia")
```

```
##  [1] "#E5B317" "#B3843D" "#98A4C2" "#D2B4AF" "#C9B5CB" "#E4C7CF" "#CBA896"
##  [8] "#E0BB68" "#F9CDB4" "#8591A4" "#D4AF57" "#987660" "#AB8D78" "#192E57"
## [15] "#7A4520" "#BAA2BD" "#A7A49E" "#EABE86" "#CCA070" "#3F372F" "#625F4B"
## [22] "#506978" "#906B32" "#F6EDF1" "#C8984A" "#ECC1A3" "#BEE1F4" "#161426"
## [29] "#7E7B7A" "#364F66" "#B8D4E2" "#668397"
```

```
t<-theme(axis.text= element_text(color =c("#E5B317" ,"#B3843D", "#98A4C2",
                                          "#D2B4AF", "#C9B5CB", "#E4C7CF",
                                          "#CBA896", "#E0BB68" ,"#F9CDB4", "#8591A4")),
         axis.title = element_text(color = "#7A4520"),legend.position='none')
```

Let's see in what season people rent more bikes.

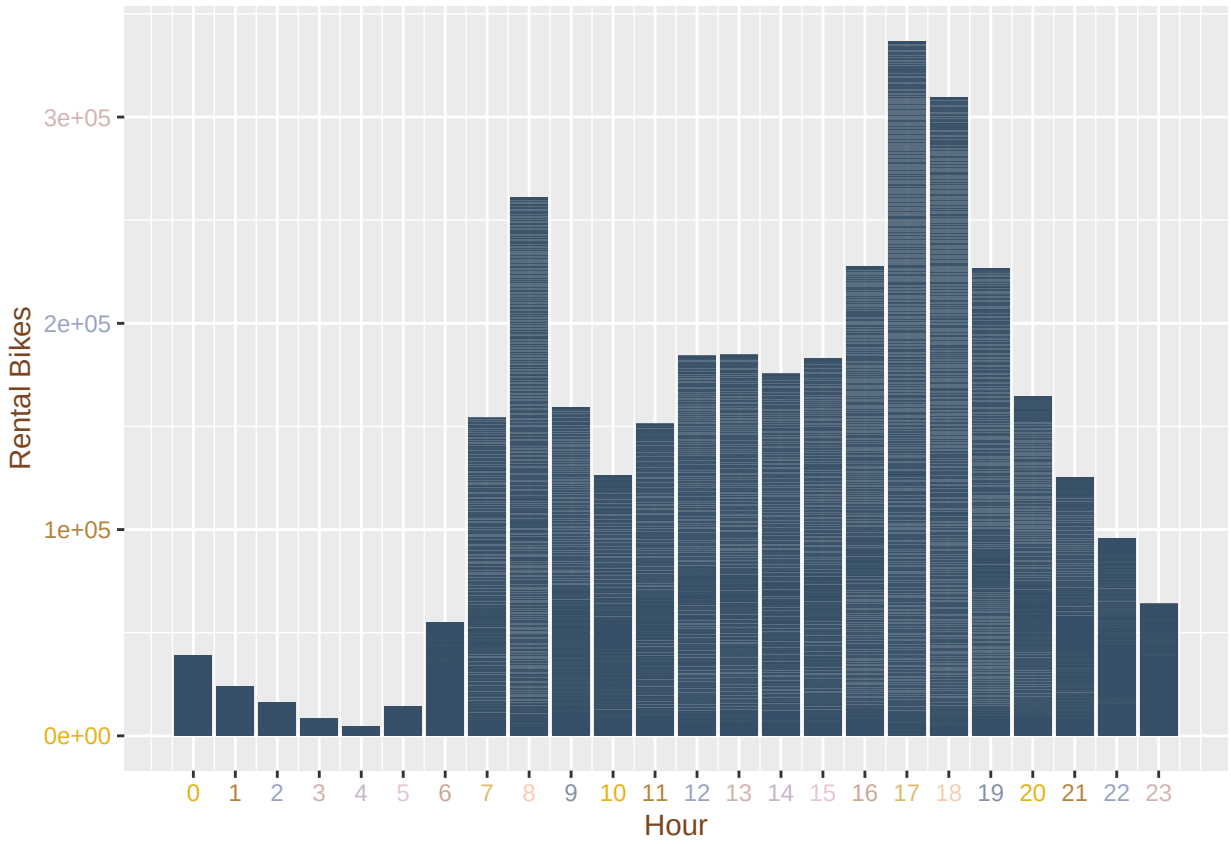
```
bike$season<-factor(bike$season)
ggplot(bike,aes(x=season,y=cnt,color=season))+
  geom_bar(stat="identity")+
  scale_x_discrete(labels =c("Winter", "Spring", "Summer", "Fall"))+
  labs(x="Seasons",y="Rental Bikes")+
  scale_color_manual(values = c("#98A4C2", "#D2B4AF" ,"#C9B5CB" ,"#E4C7CF"))+
  t
```



We see, that people rent more bike in the summer than winter.

The next let's see what time does the people ride bike most.

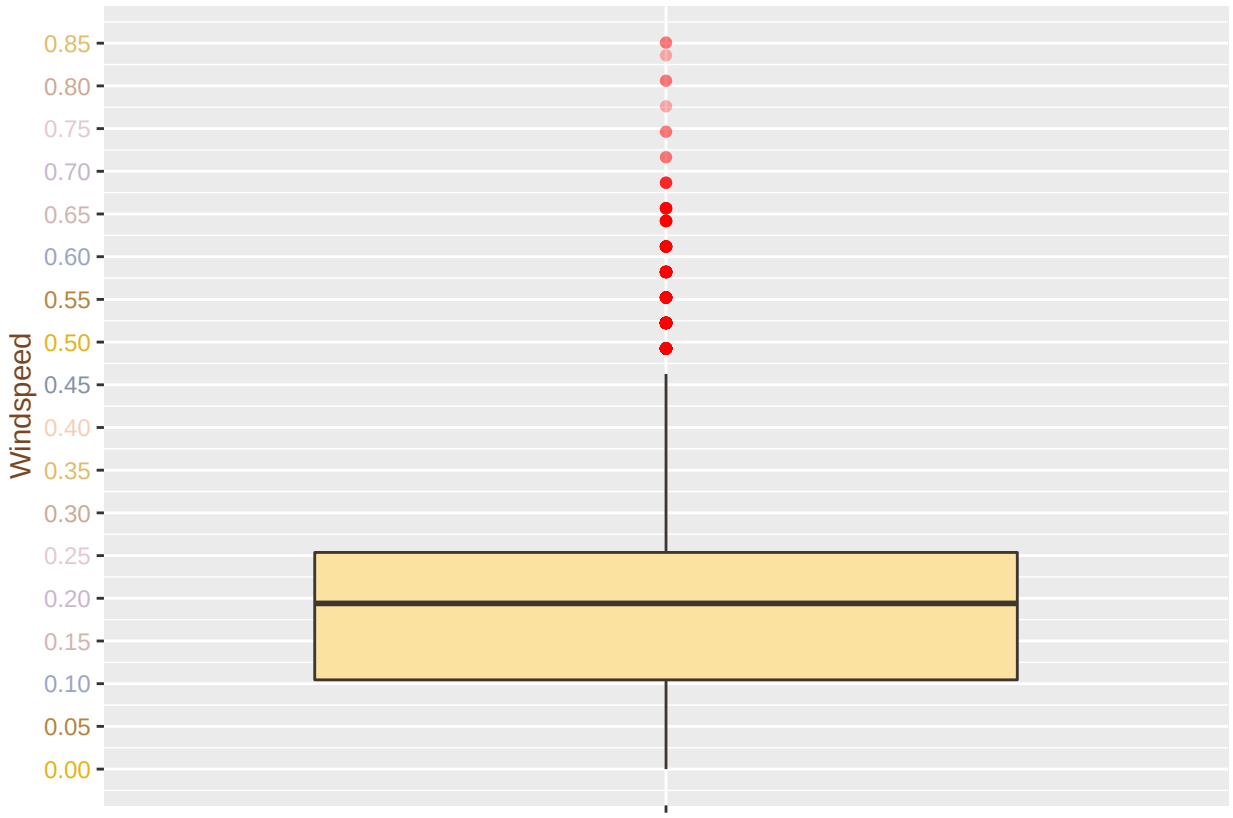
```
ggplot(bike,aes(x=hr,y=cnt))+
  geom_bar(stat="identity",fill="#364F66",alpha=0.8)+
  scale_x_continuous(breaks = 0:23)+
  labs(x="Hour",y="Rental Bikes")+
  t
```



We can conclude from this chart that people use the bikes the most whan they go home or to work. This company should have more bikes in that times to satisfy all clients.

This data we can also use to analyse some weather features.

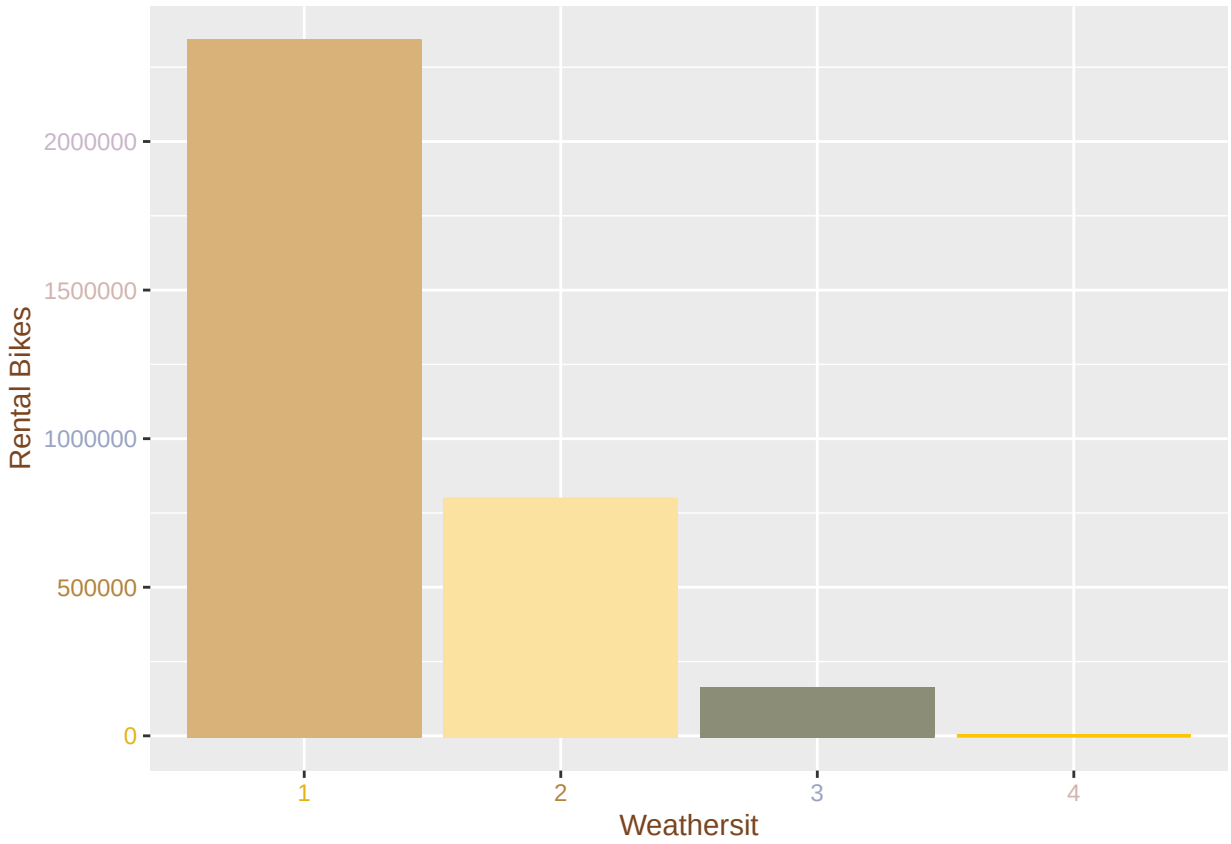
```
ggplot(bike,aes(x="",y=windspeed))+
  geom_boxplot(color="#3F372F",fill="#FCE2A1",outlier.alpha = 0.3,outlier.colour = "Red")+
  scale_y_continuous(breaks = seq(0,1,0.05))+
  ylab("Windspeed" )+xlab(NULL)+
  t
```



For example from this boxplot we see that windspeeds higher than 0.45 are outliers and median is near 0.2. 1st quartile is 0.1 and 3rd quartile is 0.25.

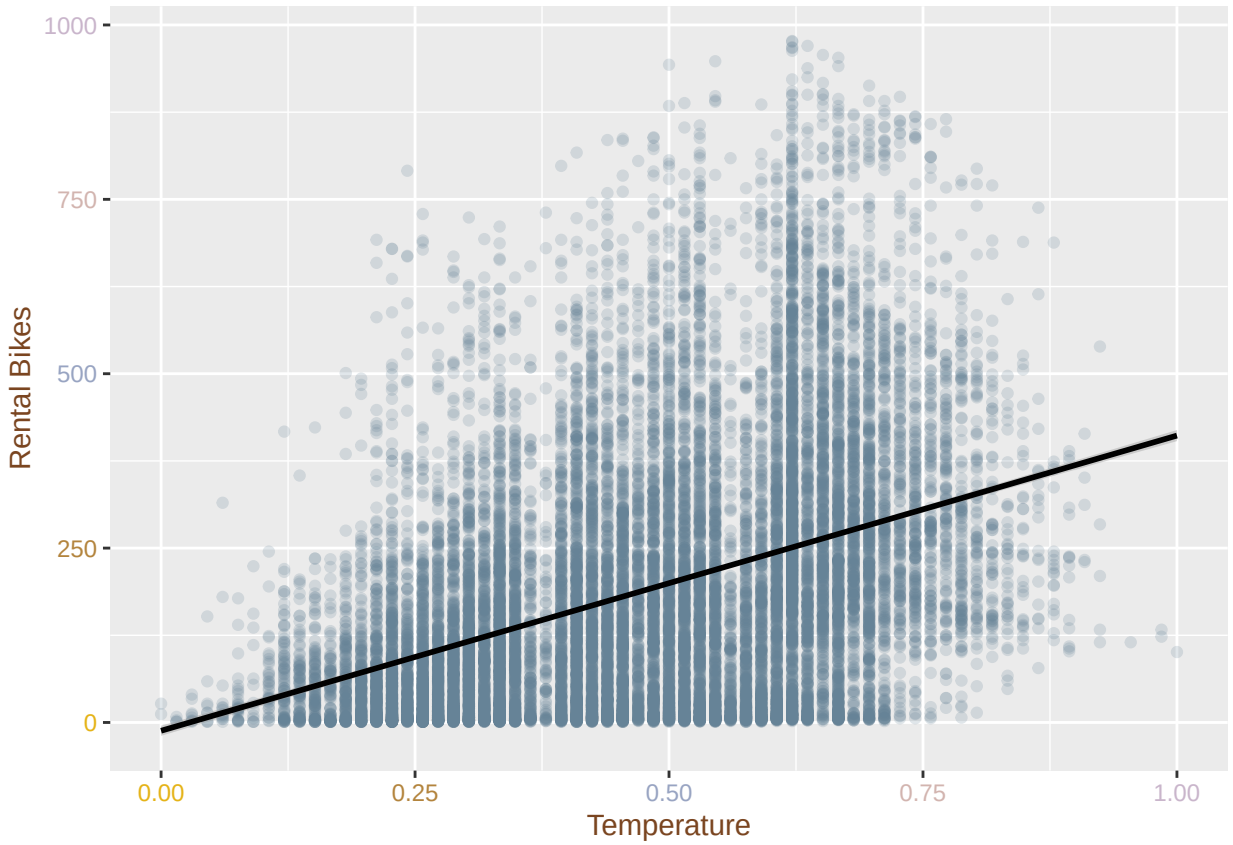
The next plot shows that people don't rent bikes in extremal weather conditions.

```
bike$weathersit<-factor(bike$weathersit)
ggplot(bike,aes(x=weathersit,y=cnt,color=weathersit))+
  geom_bar(stat = "identity")+
  scale_color_manual(values = c("#D8B279", "#FCE2A1" ,"#8C8D77" ,"#FCC604"))+
  labs(x="Weathersit",y="Rental Bikes")+
  t
```



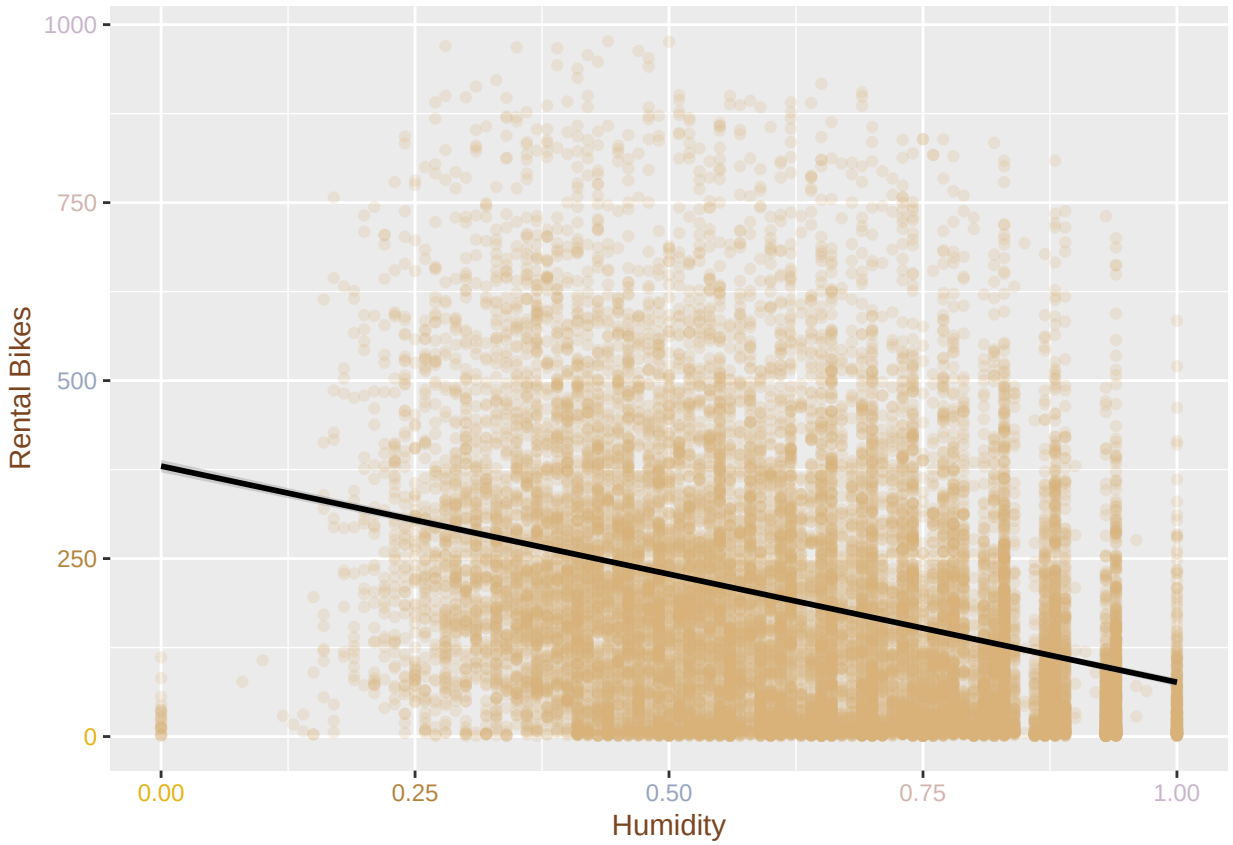
This one showes us that whan temperature rises, people rent more bikes.

```
ggplot(bike,aes(x=atemp,y=cnt))+
  geom_point(alpha=0.2,col="#668397")+
  geom_smooth(method = "lm",se=T,color=1)+
  labs(x='Temperature',y="Rental Bikes")+
  t
```



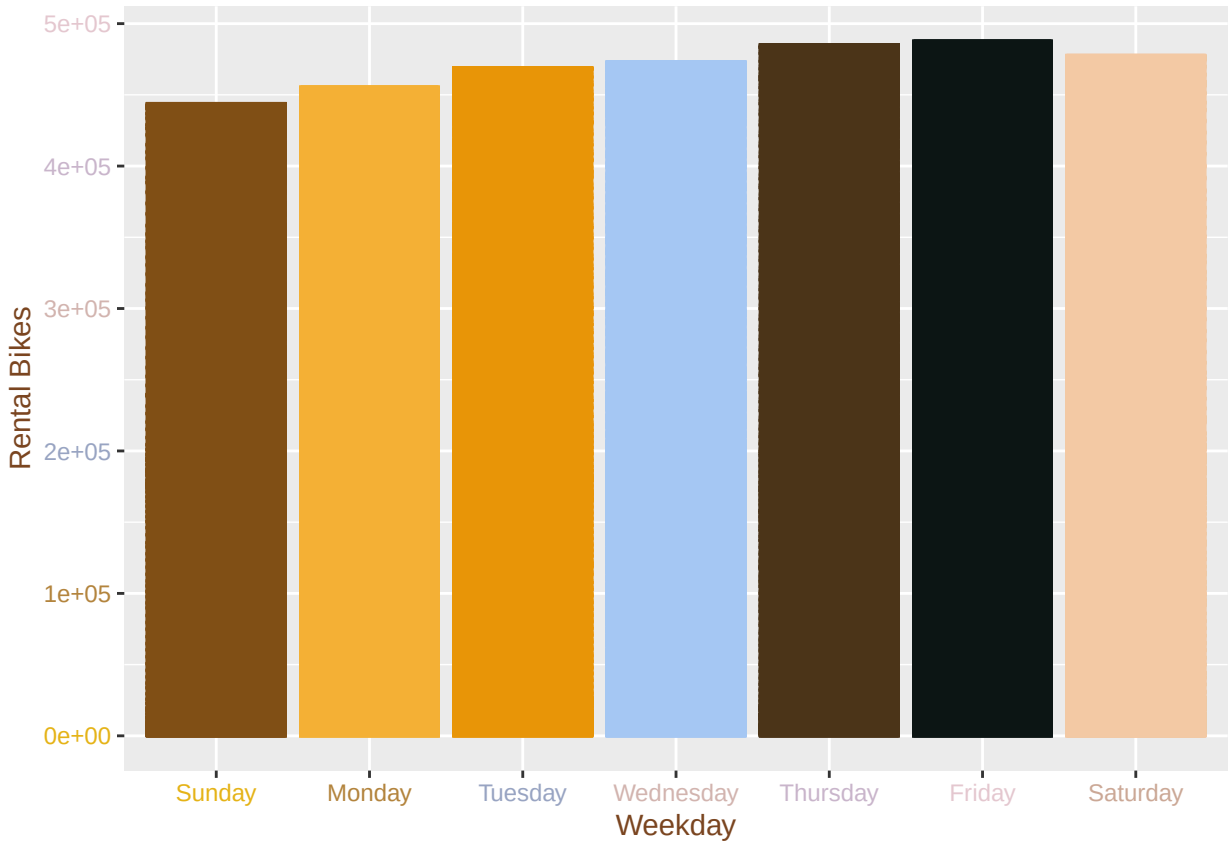
But in this plot we see that when humidity rises, people start to rent fewer bikes.

```
ggplot(bike,aes(x=hum,y=cnt))+
  geom_point(alpha=0.2,col="#D8B279")+
  geom_smooth(method = "lm",se=T,color=1)+
  labs(x="Humidity",y='Rental Bikes')+
  t
```



The next let's see if there is weekday effect.

```
bike$weekday<-factor(bike$weekday)
ggplot(bike,aes(x=weekday,y=cnt,color=weekday))+
  geom_bar(stat = "identity")+
  scale_color_manual(values=c("#804F15", "#F4B035" ,"#E89507" ,
                              "#A5C7F3", "#4B3418" ,"#0C1514", "#F3C9A4"))+
  scale_x_discrete(labels = c("Sunday","Monday", "Tuesday", "Wednesday",
                              "Thursday" ,"Friday", "Saturday"))+
  labs(x="Weekday",y="Rental Bikes")+
  t
```



We can conclude, that there is some weekday effect but it is very low.