

```

rm(list=ls(all=T))
setwd("C:/Users/Dell/Documents/Bikesharing")
getwd()
Bike_train = read.csv("day.csv", header = T, na.strings = c(" ", "", "NA"))
nrow(Bike_train)
ncol(Bike_train)
dim(Bike_train)
names(Bike_train)
#converting the normalized temperature
Bike_train$raw.temp = (Bike_train$temp*39)
head(Bike_train)
#calculating the Mean , median and standard deviation of the springer
springer <- subset(Bike_train, season == 1)$raw.temp
sp.mean <- mean(springer)
sp.median <- median(springer)
sp.sd <- sd(springer)
#calculating the Mean , median and standard deviation of the summer
summer <- subset(Bike_train, season == 2)$raw.temp
su.mean <- mean(summer)
su.median <- median(summer)
su.sd <- sd(summer)
#calculating the Mean , median and standard deviation of the fall
fall <- subset(Bike_train, season == 3)$raw.temp
fa.mean <- mean(fall)
fa.median <- median(fall)
fa.sd <- sd(fall)
#Calculating Median, Mean and Standard deviation of the winter
winter <- subset(Bike_train, season == 4)$raw.temp
wi.mean <- mean(winter)
wi.median <- median(winter)
wi.sd <- sd(winter)
#Histogram
hist(x = springer,
     main = "Temperatures in Springer",
     xlab = "Temperature in Celcius",
     ylab = "Number of Days",
     xlim = c(0, 25),
     ylim = c(0, 45))

abline(v = sp.mean, lwd = 2, lty = 1, col = "red")
text(x = 17, y = 35,
     labels = paste("Mean = ", round(mean(springer),2), sep = ""), col="red" )

abline(v = sp.median, lwd = 2, lty = 3, col = "yellow")
text(x = 6, y = 35,
     labels = paste("Median = ", round(median(springer),2), sep = ""), col="yellow" )
#Histogram of summer
hist(x = summer,
     main = "Temperatures in Summer",
     xlab = "Temperature in Celcius",
     ylab = "Number of Days",
     xlim = c(0, 35), ylim = c(0, 40)
)
abline(v = su.mean, lwd = 2, lty = 1, col = "red")
text(x = 15, y = 40,
     labels = paste("Mean = ", round(mean(summer),2), sep = ""),
     col = "red")

abline(v = su.median, lwd = 2, lty = 3, col = "green")
text(x = 31, y = 40,
     labels = paste("Median = ", round(median(summer),2), sep = ""), col = "green" )

#Histogram of fall
hist(x = fall,
     main = "Temperatures in Fall",
     xlab = "Temperature in Celcius",
     ylab = "Number of Days",
     xlim = c(15, 40), ylim = c(0, 70)
)
abline(v = fa.mean, lwd = 2, lty = 1, col = "red")
text(x = 24, y = 60,
     labels = paste("Mean = ", round(mean(fall),3), sep = ""), col = "red" )

abline(v = fa.median, lwd = 2, lty = 3, col = "blue")
text(x = 35, y = 60,
     labels = paste("Median = ", round(median(fall),3), sep = ""), col = "blue" )
#Histogram of winter
hist(x = winter,
     main = "Temperatures in Winter",
     xlab = "Temperature in Celcius",
     ylab = "Number of Days",
     xlim = c(0, 30), ylim = c(0, 40)
)

abline(v = wi.mean, lwd = 2, lty = 1, col = "red")

```

```

text(x = 23, y = 40,
     labels = paste("Mean = ", round(mean(winter),2), sep = ""), col = "red" )

abline(v = wi.median, lwd = 2, lty = 3, col ="blue")
text(x = 10, y = 40,
     labels = paste("Median = ", round(median(winter),2), sep = ""), col ="blue")
#check dataset
is.integer(Bike_train)
#check null
is.null(Bike_train)
#new column
hist(x = winter,
     main = "Temperatures in Winter",
     xlab = "Temperature in Celcius",
     ylab = "Number of Days",
     xlim = c(0, 30), ylim = c(0, 40)
)

abline(v = wi.mean, lwd = 2, lty = 1, col = "red")
text(x = 23, y = 40,
     labels = paste("Mean = ", round(mean(winter),2), sep = ""), col = "red" )

abline(v = wi.median, lwd = 2, lty = 3, col ="blue")
text(x = 10, y = 40,
     labels = paste("Median = ", round(median(winter),2), sep = ""), col ="blue" )
hist(x = winter,
     main = "Temperatures in Winter",
     xlab = "Temperature in Celcius",
     ylab = "Number of Days",
     xlim = c(0, 30), ylim = c(0, 40)
)

abline(v = wi.mean, lwd = 2, lty = 1, col = "red")
text(x = 23, y = 40,
     labels = paste("Mean = ", round(mean(winter),2), sep = ""), col = "red" )

abline(v = wi.median, lwd = 2, lty = 3, col ="blue")
text(x = 10, y = 40,
     labels = paste("Median = ", round(median(winter),2), sep = ""), col ="blue" )
#create new column
Bike_train$raw.atep <- (Bike_train$atemp * 50)
Bike_train$raw.mean.temp.atep <- (Bike_train$raw.temp + day$raw.atep)/2
head(Bike_train)
#correlation test
cor.temp <- cor.test(x = Bike_train$raw.temp,
                     y = Bike_train$cnt)

cor.temp
Temperature <- Bike_train$raw.temp
Amount.Rentals <- Bike_train$cnt
#Correlation in mean.temp.atep and total count of bike rental
Bike_train$raw.mean.temp.atep <- (Bike_train$raw.temp + Bike_train$raw.atep)/2
cor.mean.temp.atep <- cor.test(x = Bike_train$raw.mean.temp.atep,
                              y = Bike_train$cnt)

cor.mean.temp.atep
#correlation plot
par(mfrow=c(2,2))

plot(x = Temperature, y = Amount.Rentals, main = "Correlation", col = "red")
abline(lm(Amount.Rentals ~ Temperature), col = "blue")
legend("topleft", legend = paste("cor = ", round(cor(Temperature, Amount.Rentals), 2), sep = ""),lty = 1, col =
"blue")

plot(x = Feeled.Temperature, y = Amount.Rentals, main = "Correlation", col = "blue")
abline(lm(Amount.Rentals ~ Feeled.Temperature), col = "red")
legend("topleft", legend = paste("cor = ", round(cor(Feeled.Temperature, Amount.Rentals), 2), sep = ""),lty = 1, col =
"red")

plot(x = Feeled.Raw.Temperature, y = Amount.Rentals, main = "Correlation", col = "green")
abline(lm(Amount.Rentals ~ Feeled.Raw.Temperature), col = "orange")
legend("topleft", legend = paste("cor = ", round(cor(Temperature, Amount.Rentals), 2), sep = ""),lty = 1, col =
"orange")

plot(x = 1, y = 1, xlab = "Temperature", ylab = "Amount of rentals", xlim = c(0, 40), ylim = c(0, 10000), main =
"Three correlations combined")

points(Feeled.Raw.Temperature, Amount.Rentals, pch = 8, col = "green")
points(Temperature, Amount.Rentals, pch = 8, col = "red")
points(Feeled.Temperature, Amount.Rentals, pch = 8, col = "blue")
#difference feeled temp to real temp
test.result.1 <- t.test(x = Bike_train$raw.temp, y = Bike_train$raw.atep, alternative = "two.sided")
test.result.1
hist(Bike_train$raw.temp, yaxt = "n", xaxt = "n", xlab = "",
     ylab = "", main = "Two Sample t-test", xlim = c(5, 40), col = rgb(0, 0, 1, alpha = .1))

```

```

text(x = 13, y = 140, paste("Mean real Temp.\n", round(mean(day$raw.temp), 2), sep = ""), col = "blue")
abline(v = mean(day$raw.temp), lty = 1,
       col = rgb(0, 0, 1, alpha = 1), lwd = 4)

par(new = T)
hist(Bike_train$raw.atemp, yaxt = "n", xaxt = "n", xlab = "",
     ylab = "", main = "", xlim = c(5, 40), col = rgb(1, 0, 0, alpha = .1))

abline(v = mean(Bike_train$raw.atemp), lty = 1,
       col = rgb(1, 0, 0, alpha = 1), lwd = 4)
text(x = 32, y = 131, paste("Mean feelled Temp.\n", round(mean(Bike_train$raw.atemp), 2), sep = ""), col = "red")

mtext(text = "Alternative Hypothesis is confirmed true difference in means is not equal to 0", line = 0, side = 3)
#Two Sample t-test across the seasons

temp.spring <- subset(Bike_train, subset = season == "1")$raw.temp
atemp.spring <- subset(Bike_train, subset = season == "1")$raw.atemp
test.result.spring <- t.test(x = temp.spring, y = atemp.spring, alternative = "two.sided")
test.result.spring
# two-sample t-test for summer.

temp.summer <- subset(Bike_train, subset = season == "2")$raw.temp
atemp.summer <- subset(Bike_train, subset = season == "2")$raw.atemp
test.result.summer <- t.test(x = temp.summer, y = atemp.summer, alternative = "two.sided")
test.result.summer

#two-sample t-test for fall.

temp.fall <- subset(Bike_train, subset = season == "3")$raw.temp
atemp.fall <- subset(Bike_train, subset = season == "3")$raw.atemp
test.result.fall <- t.test(x = temp.fall, y = atemp.fall, alternative = "two.sided")
test.result.fall
#two-sample t-test for winter.

temp.winter <- subset(Bike_train, subset = season == "4")$raw.temp
atemp.winter <- subset(Bike_train, subset = season == "4")$raw.atemp
test.result.winter <- t.test(x = temp.winter, y = atemp.winter, alternative = "two.sided")
test.result.winter
#Plotting the association:
plot(x = 1, y = 1, xlab = "Temperature in Celcius", ylab = "Bike rentals", type = "n", main = "Association between
temperature and bike rentals",
     xlim = c(0, 40), ylim = c(0, 6000))

# min and max for the x-axis and y-axis:
min(Bike_train$raw.temp)
max(Bike_train$raw.temp)
min(Bike_train$casual)
max(Bike_train$casual)
min(Bike_train$registered)
max(Bike_train$registered)
#points to the plot
Bike_train$raw.temp <- (Bike_train$temp*39)
points(Bike_train$raw.temp, Bike_train$casual, pch = 16, col = "red")
points(Bike_train$raw.temp, Bike_train$registered, pch = 16, col = "skyblue")

# legend to the plot
legend("topleft", legend = c("casual", "registered"), col = c("red", "skyblue"), pch = c(16, 16), bg = "green")

# correlation between raw.temp vs registered users between raw.temp and causal users
cor.reg <- cor.test(x = Bike_train$raw.temp, y = Bike_train$registered)
cor.reg
day$raw.temp <- (day$temp*41)
points(day$raw.temp, day$casual, pch = 16, col = "red")
points(day$raw.temp, day$registered, pch = 16, col = "skyblue")

# Adding a legend to the plot
legend("topleft", legend = c("casual", "registered"), col = c("red", "skyblue"), pch = c(16, 16), bg = "white")

cor.reg <- cor.test(x = Bike_train$raw.temp, y = Bike_train$registered)
cor.reg
cor.cas <- cor.test(x = Bike_train$raw.temp,
                   y = Bike_train$casual)
cor.cas
#adding Correlation line and the correlation value to the plot
abline(lm(Bike_train$registered ~ Bike_train$raw.temp), lty = 6, col = "blue")

abline(lm(Bike_train$casual ~ Bike_train$raw.temp), lty = 6, col = "orange")

reg <- paste("cor = ", round(cor(Bike_train$registered, Bike_train$raw.temp), 2), sep = "")
cas <- paste("cor = ", round(cor(Bike_train$casual, Bike_train$raw.temp), 2), sep = "")

```

```

legend("left",legend = c(cas, reg) , col = c('orange', 'blue'),pch = c(16, 16), bg = "white")
#converting weather
lookup <- data.frame("numbers"=c("1","2","3","4"),
                     "weather"=c("lousy","wet", "cloudy", "nice")
)

Bike_train<- merge(x= Bike_train,
                  y= lookup,
                  by.x="weathersit",
                  by.y="numbers",
)

head(Bike_train)
#Using linear regression

total.rentals.lm <- lm(cnt ~ holiday + weather, data = Bike_train)
summary (total.rentals.lm)
# Using a anova
anv.weather <- anova (total.rentals.lm)
anv.weather
anv.weather$`F value`
anv.weather$`Pr(>F)`
#test
weather.aov <- aov(cnt ~ weather, data = Bike_train)
summary(weather.aov)
. # The mean temperature, humidity, windspeed and total rentals per months?
  # Converting month with "merge"

lookup.month<- data.frame("mnth" = c(1:12),
                          "mnth.name" = c("01Jan", "02Feb", "03March", "04April", "05May", "06June", "07July",
"08Aug", "09Sept", "10Oct", "11Nov", "12Dec"), stringsAsFactors = FALSE)

Bike_train <- merge(x=Bike_train, y= lookup.month, by = 'mnth')

# Convert the nomalized windspeed and humidity
Bike_train$raw.windspeed <- (Bike_train$windspeed*67)
Bike_train$raw.hum <- (Bike_train$hum * 100)
head(Bike_train)
require(dplyr)
month.agg <- Bike_train %>% group_by(mnth.name) %>% summarise(
  mean.temp = mean(raw.temp),
  mean.hum = mean(raw.hum),
  mean.windspeed = mean(raw.windspeed),
  mean.rentals = mean(cnt))
month.agg

month.agg
#mean associated with month
par(mfrow=c(2,2))
barplot(height = month.agg$mean.rentals,
        names.arg = month.agg$mnth.name ,col = "red", main = "Mean rentals" )

barplot(height = month.agg$mean.windspeed,
        names.arg = month.agg$mnth.name,col = "blue", main = "Mean Windspeed (km/h)" )

barplot(height = month.agg$mean.hum,
        names.arg = month.agg$mnth.name,col = "green", main = "Mean Humidity" )

barplot(height = month.agg$mean.temp,
        names.arg = month.agg$mnth.name,col = "skyblue", main = "Mean Temperature" )
max(Bike_train$raw.temp)
#fine weather for biking

biking.Bike_train <- function (temp.thresh, windspeed.thresh, weathersit.thresh)
{result <- with (Bike_train, raw.temp > temp.thresh &
                raw.windspeed < windspeed.thresh &
                weathersit < weathersit.thresh)

return(result)}

mean(biking.Bike_train(5, 40, 3))
mean(biking.Bike_train(10, 20, 2))

```