Q3 天氣是否影響乘客乘坐出租車？解釋如何。

library(ggplot2)

library(dplyr)

#載入資料

path\_sample <- "C:\\Users\\Yuru\\Documents\\YURU\\FCU\\課\\1061\\1061 R語言資料科學基石(資電綜班)[4441]\\專題\\dw\_exercise1\\sample.csv"

sample <- read.csv(pathsample,TRUE,",")

sample <-filter(sample,trip\_distance>0)

path\_temperature <- "C:\\Users\\Yuru\\Documents\\YURU\\FCU\\課\\1061\\1061 R語言資料科學基石(資電綜班)[4441]\\專題\\dw\_exercise1\\temperature.txt"

temperature <- read.csv(path\_temperature,FALSE,",")

temperature <- sapply(temperature,"[[",1)

names(temperature)<-NULL

path\_humidity <- "C:\\Users\\Yuru\\Documents\\YURU\\FCU\\課\\1061\\1061 R語言資料科學基石(資電綜班)[4441]\\專題\\dw\_exercise1\\humidity.txt"

humidity <- read.csv(path\_humidity,FALSE,",")

humidity <- sapply(humidity,"[[",1)

names(humidity)<-NULL

path\_precipitation <- "C:\\Users\\Yuru\\Documents\\YURU\\FCU\\課\\1061\\1061 R語言資料科學基石(資電綜班)[4441]\\專題\\dw\_exercise1\\precipitation.txt"

precipitation <- read.csv(path\_precipitation,FALSE,",")

precipitation <- sapply(precipitation,"[[",1)

names(precipitation)<-NULL

#讀出計程車乘車日期

info\_day <- dplyr::select(sample,tpep\_pickup\_datetime,tpep\_dropoff\_datetime) %>%

mutate(day=substring(tpep\_pickup\_datetime,9,10),hot,wet,rain) %>%

select(day)

class(info\_day$day) <- "integer"

day <- info\_day$day

hot <- temperature[day]>mean(temperature)

wet <- humidity [day]>mean(humidity)

rain <- precipitation[day]>mean(precipitation)

info\_day$hot <- hot

info\_day$wet <- wet

info\_day$rain <- rain

#計算總和

sum(hot) #72781

sum(!hot) #49686

sum(wet) #57551

sum(!wet) #64916

sum(rain) #35815

sum(!rain) #86652

ggplot(info\_day,aes(x=hot)) +

geom\_bar() +

ggtitle("Weather(temperature) affects customers to take taxi?") +

theme\_bw() +

theme(plot.title=element\_text(hjust=0.5,size=18))

ggplot(info\_day,aes(x=wet)) +

geom\_bar() +

ggtitle("Weather(humidity) affects customers to take taxi?") +

theme\_bw() +

theme(plot.title=element\_text(hjust=0.5,size=18))

ggplot(info\_day,aes(x=rain)) +

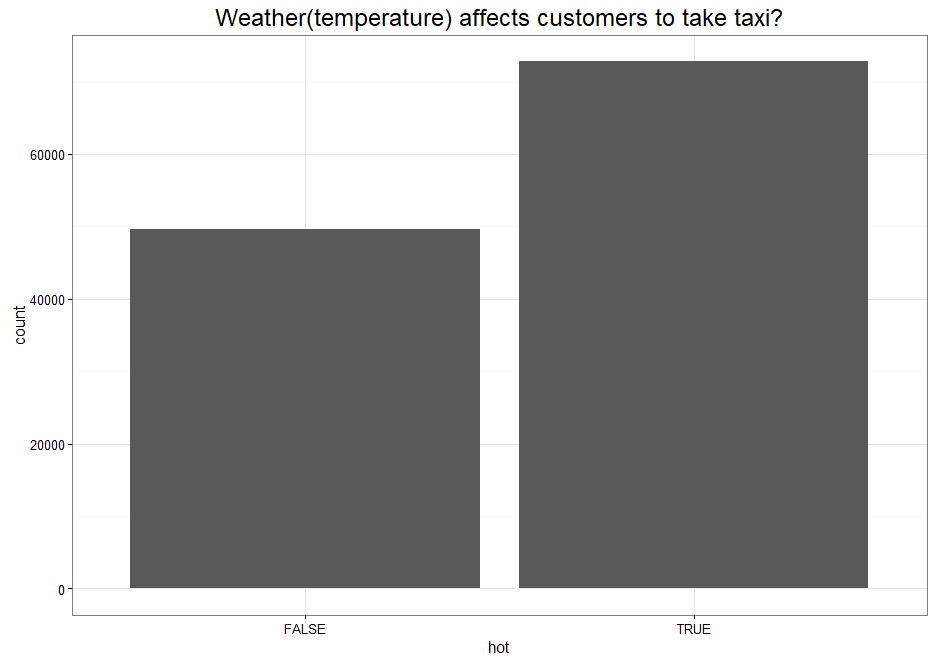
geom\_bar() +

ggtitle("Weather(precipitation) affects customers to take taxi?") +

theme\_bw() +

theme(plot.title=element\_text(hjust=0.5,size=18))

View(info\_day)

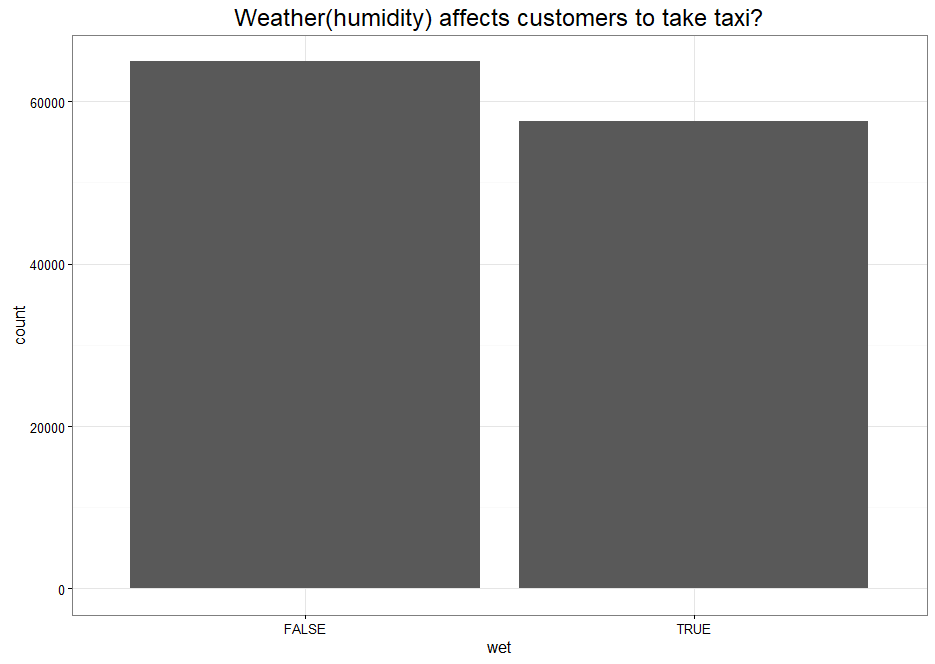


以平均溫度(21.86667)作為分界，分為冷熱分界

在天氣熱時搭車總數: 72781

在天氣冷時搭車總數: 49686

天氣冷時較少人外出，較少人搭車。

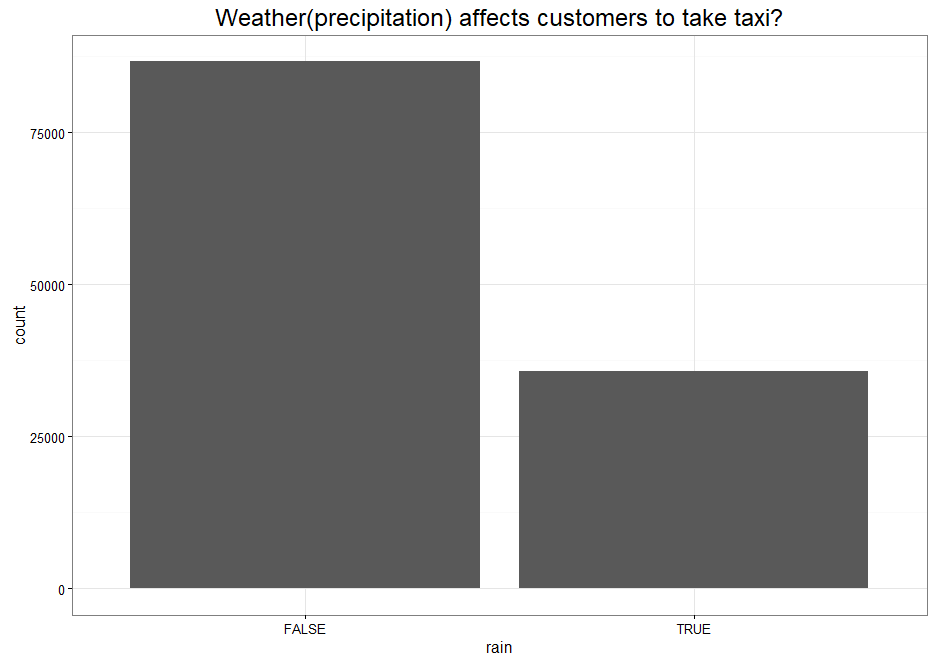


以平均濕度(65.03333)作為分界，分為乾濕分界

在天氣乾時搭車總數:57551

在天氣濕時搭車總數: 64916

天氣乾溼較不影響搭車總數。



以平均降雨量(4.022)作為分界，分為有無降雨分界

雨量大小的測量標準:

小雨（每小時降雨0.5mm以下）

中雨（每小時降雨0.5mm ~ 4mm）

豪雨、大雨（每小時降雨4mm以上）

當日總雨量4mm以下，雨量極少，相當於無下雨

在有下雨時搭車總數: 35815

在無下雨時搭車總數: 86652

因為當地有地鐵，下雨時可能搭乘較方便且不淋雨的地鐵來做為交通工具，有無下雨影響乘車意願甚大。

總結:

天氣會影響乘客乘坐計程車的意願。

Q4 長途旅行意味著更多小費？

library(ggplot2)

library(dplyr)

path\_sample <- "C:\\Users\\Yuru\\Documents\\YURU\\FCU\\課\\1061\\1061 R語言資料科學基石(資電綜班)[4441]\\專題\\dw\_exercise1\\sample.csv"

sample <- read.csv(pathsample,TRUE,",")

info\_tip <- select(sample,trip\_distance,tip\_amount) %>%

filter(trip\_distance>0,tip\_amount>0)%>%

filter(tip\_amount<50)

cor(info\_tip)

ggplot(info\_tip,aes(x=trip\_distance,y=tip\_amount)) +

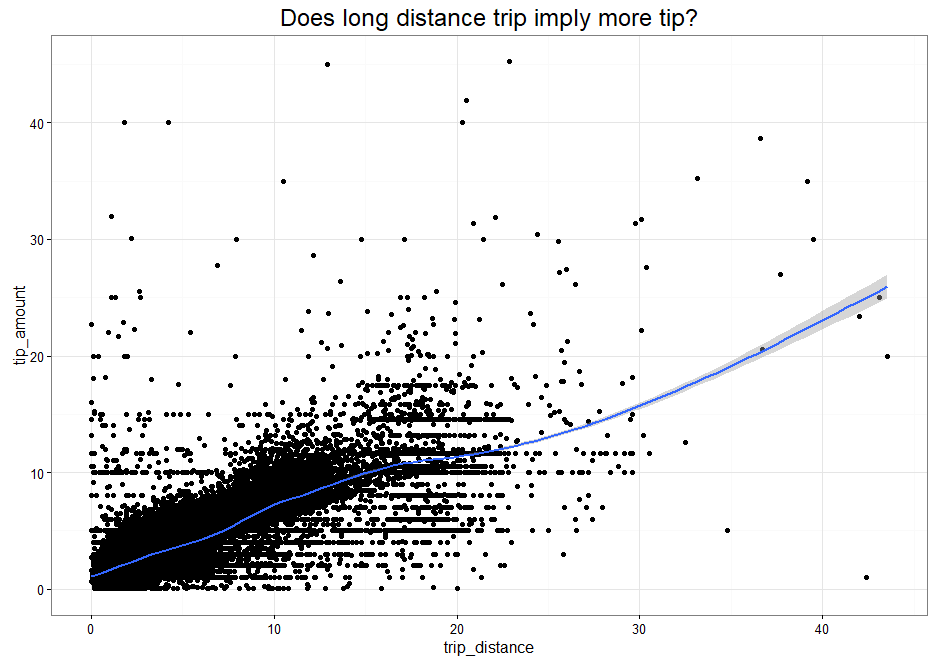
geom\_point() +

geom\_smooth() +

ggtitle("Does long distance trip imply more tip?") +

theme\_bw() +

theme(plot.title=element\_text(hjust=0.5,size=18))



扣除無小費者，可能有缺漏紀錄資料或沒有習慣給小費的例外情形，只以有給小費做為考慮

扣除極端值，小費金額大多集中在20元之內，較長途則在40元內，超過50元的小費可能有其他原因，所以給予更多小費

扣除以上兩項，路途與小費的金額成正比關係，相關係數為0.8265739為高度相關(0.7以上)。

總結:

長途旅行會有更多的小費。