# 先確定下面四個套件已經安裝並在user library有勾選

## install.packages("qqplot2", dependencies = TRUE,

## repos="http://cran.us.r-project.org")

## install.packages("formatR", dependencies = TRUE,

## repos="http://cran.us.r-project.org")

## install.packages("highlight", dependencies = TRUE,

## repos="http://cran.us.r-project.org")

## install.packages("xtable", dependencies = TRUE,

## repos="http://cran.us.r-project.org")

require(formatR)

require(highlight)

require(xtable)

require(ggplot2)

## library(formatR)

## library(highlight)

## library(xtable)

## library(ggplot)

## library(xtable)

# 設定工作目錄

setwd("C:/RData")

# 確定EXCEL檔avgpm25.csv在目錄C:/RData下

# 將EXCEL檔案讀入R, colClasses行變數的類型

pollution <- read.csv("C:/RData/avgpm25.csv",

 colClasses=c("numeric", "character",

 "factor","numeric", "numeric"))

# 印出空氣汙染這個data frame 的第一部分(前6筆)

head(pollution)



# 細懸浮微粒（PM2.5，直徑≦2.5 微米）

# 印出細懸浮微粒的5個數字的摘要(5-number summary)加平均數

# pollution$pm25 代表在data frame pollution中的pm25變數, 用$號連結

summary(pollution$pm25)



# 畫細懸浮微粒的盒狀圖, 顏色用藍色

boxplot(pollution$pm25, col = "blue")



# 分配大致對稱, 有一些outliers

# 畫細懸浮微粒的直方圖, 顏色用綠色, y軸是次數

hist(pollution$pm25, col = "green")



# 畫細懸浮微粒的直方圖, 顏色用綠色, y軸是次數

hist(pollution$pm25, col = "green")

rug(pollution$pm25) # 將數據加到圖下方(將數據用地毯表示)



# 畫細懸浮微粒的直方圖, 顏色用綠色, y軸是次數, 分成100格

hist(pollution$pm25, col = "green", breaks = 100)

rug(pollution$pm25)



# 畫細懸浮微粒的盒狀圖, 顏色用藍色

boxplot(pollution$pm25, col = "blue")

abline(h = 12) # 在位置12加畫一條水平線



# 畫細懸浮微粒的直方圖, 顏色用綠色

hist(pollution$pm25, col = "green")

abline(v = 12, lwd = 2) # 在位置12加畫一條垂直線, 線條寬度為2

abline(v = median(pollution$pm25), col = "magenta", lwd = 4) # 在中位數位置加畫一條垂直線, 線條寬度為4, 顏色為品紅



# 建立變數區域的次數分配表

table(pollution$region)



# 畫區域的次數的長條圖, 顏色是麥色, 圖標題為"Number of Counties in Each Region"

barplot(table(pollution$region), col = "wheat", main = "Number of Counties in Each Region")



# 畫按區域分組的 細懸浮微粒的盒狀圖, 顏色用紅色

boxplot(pm25 ~ region, data = pollution, col = "red")



# 由圖可知空氣中的細懸浮微粒高於西區，東區空氣汙染較嚴重

# 設定繪圖參數, mfrow = c(1, 2)代表圖將放在1列2行的矩陣, mar = c(4, 4, 2, 1)代表圖的下左上右的邊寬

par(mfrow = c(1, 2), mar = c(4, 4, 2, 1))

hist(subset(pollution, region == "east")$pm25, col = "green")

hist(subset(pollution, region == "west")$pm25, col = "green")



# 左邊是東區的細懸浮微粒的直方圖, 右邊是東區的細懸浮微粒的直方圖

# 左邊東區的直方圖可知細懸浮微粒的分佈大致呈現對稱, 右邊西區的直方圖可知細懸浮微粒的分佈大致呈現右偏

# 選出區域為east的細懸浮微粒的資料可利用subset指令

subset(pollution, region == "east")$pm25



## # with 是指Evaluate an Expression in a Data Environment

## # 畫散佈圖, x軸是緯度, y軸是細懸浮微粒

with(pollution, plot(latitude, pm25))



abline(h = 12, lwd = 2, lty = 2) #在位置12加畫一條水平線, 線條寬度為2, 線條類型為2, 虛線



## # 由散佈圖可看出緯度與細懸浮微粒沒有相關

## # 畫散佈圖, x軸是緯度, y軸是細懸浮微粒, 亦可用下面指令

plot(pollution$latitude, pollution$pm25, xlab="latitude", ylab="pm25")



## # 畫散佈圖, x軸是緯度, y軸是細懸浮微粒, 按區域分顏色, 東區是黑色, 西區是紅色

with(pollution, plot(latitude, pm25, col = region))

abline(h = 12, lwd = 2, lty = 2) #在位置12加畫一條水平線, 線條寬度為2, 線條類型為2, 虛線



# 西區細懸浮微粒較分散, 東區細懸浮微粒較集中

# 設定繪圖參數, mfrow = c(1, 2)代表圖將放在1列2行的矩陣, mar = c(4, 4, 2, 1)代表圖的下左上右的邊寬

par(mfrow = c(1, 2), mar = c(4, 4, 2, 1))

## # 畫散佈圖, 區域為西區, x軸是緯度, y軸是細懸浮微粒

with(subset(pollution, region == "west"), plot(latitude, pm25, main = "West"))

## # 畫散佈圖, 區域為東區, x軸是緯度, y軸是細懸浮微粒

with(subset(pollution, region == "east"), plot(latitude, pm25, main = "East"))



# 設定繪圖參數, mfrow = c(1, 2)代表圖將放在1列2行的矩陣, mar = c(4, 4, 2, 1)代表圖的下左上右的邊寬

par(mfrow = c(1, 2), mar = c(4, 4, 2, 1))

## # 畫散佈圖, 區域為西區, x軸是緯度, y軸是細懸浮微粒, y軸範圍為1到20

with(subset(pollution, region == "west"), plot(latitude, pm25, main = "West", ylim=c(1,20)))

## # 畫散佈圖, 區域為東區, x軸是緯度, y軸是細懸浮微粒, y軸範圍為1到20

with(subset(pollution, region == "east"), plot(latitude, pm25, main = "East" , ylim=c(1,20)))



# y軸範圍一樣, 才可看出細懸浮微粒的分散程度

# 西區緯度與細懸浮微粒無關, 東區緯度與細懸浮微粒有非線性相關