```
########################
# Set Working Directory #
########################

# Get your current working directory #
getwd() # Re-check the path for the working directory

# Change your current working directory #
setwd("/Users/huangweiting/coding/INTRODUCTION TO SCIENTIFIC COMPUTING SOFTWARE
/C3_ClassData")
getwd() # Re-check the path for the working directory

###########################
# Import an example dataset #
###########################

dataset <- read.csv("C3_HW.csv", fileEncoding ="utf-8")
#dataset <- read.table("C3.csv",header = TRUE, sep = ",")  #Import dataset


#str(dataset)        #Check the variable format
#View(dataset)           #Check Dataset
#dim(dataset)            #Check Dataset (how many observations and variables)
#colnames(dataset)       #Check variable names
#summary(dataset)        #Get the summary statistics about the object

####################
# generate a boxplot #
####################

install.packages("lattice")
library("lattice")

#boxplot(formula= Rainfall~SeasonNum ,data =dataset, xlab = "Season", ylab =
"Rainfall(mm)", col ="red")

#boxplot(formula= Rainfall~Town_name,data =dataset, xlab = "Town", ylab =
"Rainfall(mm)", col ="blue")

dataset2 <- na.omit(dataset) #刪掉有缺失的數據

#######################Question#####################
#Q1
#2009-2011年三年間全台溫度是否為常態分布，並描述資料分布狀況
#常態分布需附上使用的分析方法、p值
##Skewness & Kurtosis ##
install.packages("psych")
library(psych)
skew(dataset2$temperature)    #Skewness
kurtosi(dataset2$temperature)  #Kurtosis
#The Shapiro-Wilk Normality Test
#shapiro.test(dataset2$Rainfall)    #小樣本
install.packages("nortest")
library(nortest)
#Lilliefors(Kolmogorov-Smirnov) Normality Test
lillie.test(dataset2$temperature) #大樣本
#需提到資料分布狀況之偏度以及峰度，可畫圖輔助，例：直方圖
#install.packages("lattice")
# How to load a package #
library("lattice")
histogram(x= ~temperature,data =dataset2, xlab ="temperature(Degree Celsius)" ,ylab =
"count" ,type = "count",main="Temperature distribution in Taiwan during 2009-2011")

#Q2 ok
boxplot(formula= temperature~season ,data =dataset2, xlab = "season", ylab =
"temperature(Degree Celsius)", col ="red",main="Box plot of temperature in four seasons
```

```
in Taiwan from 2009 to 2011")
```

#Q3
#請解答2009-2011年三年間新北市溫度是否為常態分布，並描述資料分布狀況

```
dataset_select <- subset(dataset2,city=="新北市")
lillie.test(dataset_select$temperature)
skew(dataset_select$temperature)
kurtosi(dataset_select$temperature)
histogram(x= ~temperature,data =dataset_select, xlab ="temperature(Degree Celsius)"
,ylab = "count" ,type = "count",main="Temperature distribution in New Taipei City during
2009-2011")
```

#Q4
#依據年份，畫出2009-2011年每一年新北市溫度之盒形圖(boxplot) 。

#x軸為年份，y軸為溫度，三年份結果繪製在同一張圖)

```
#"Box plot of annual temperature in New Taipei City from 2009 to 2011"
boxplot(formula= temperature~year ,data =dataset_select, xlab = "year", ylab =
"temperature(Degree Celsius)", col ="red",main="Box plot of annual temperature in New
Taipei City from 2009 to 2011")
```