**M9\_Exercise\_LinePlotLinearFit\_Abhishek\_Jain**

The entire R code used when creating the line plot in (1).

*# Loading the datasets*

summerOlympics <- read.csv("summer.csv")

winterOlympics <- read.csv("winter.csv")

cherryBlossoms <- read.csv("cherry.csv")

temperature <- read.csv("washtemp.csv", header = TRUE, sep = ' ')

summerOlympics$ev <- paste(summerOlympics$Discipline, summerOlympics$Event,

summerOlympics$Gender, sep="\_")

summerOlympicsEvents <- summerOlympics[c("Year", "ev")]

summerOlympicsEventsUnique <- unique(summerOlympicsEvents)

summerOlympicsEventsCount <- table(summerOlympicsEventsUnique$Year)

*# Line plot of number of events each year*

plot(summerOlympicsEventsCount, xlab = "Year", ylab = "Number of Events",

main = "Number of Events at Summer Olympics through years", type="l")

Screenshot of the line plot created in (1).

Chart, line chart

Description automatically generated

The entire R code used when creating the data frame in (2), scatter plot in (3), line in (4), and prediction in (5).

summerOlympicsSubset <- as.data.frame(summerOlympicsEventsCount[11:27])

names(summerOlympicsSubset) = c("Year", "Events")

summerOlympicsSubset$Year <- as.numeric(as.character(summerOlympicsSubset$Year))

*# Scatter plot of data from 1950 onwards*

plot(summerOlympicsSubset, xlab = "Year", ylab = "Number of Events",

main="Number of Events at Summer Olympics from 1950 onwards")

*# Line fit*

linefit1 <- lm(Events~Year, data = summerOlympicsSubset)

abline(linefit1, col = "red", lwd = 4)

*# Predicting number of events in 2040*

predict(linefit1, list(Year=2040))

Screenshot of the scatter plot created in (3) with the line created in (4).

Chart, scatter chart

Description automatically generated

The prediction (answer) made in (5).

A picture containing text

Description automatically generated

The entire R code used when creating the line plot in (6).

winterOlympics$ev <- paste(winterOlympics$Discipline, winterOlympics$Event,

winterOlympics$Gender, sep="\_")

winterOlympicsEvents <- winterOlympics[c("Year", "ev")]

winterOlympicsEventsUnique <- unique(winterOlympicsEvents)

winterOlympicsEventsCount <- table(winterOlympicsEventsUnique$Year)

*# Line plot of number of events each year*

plot(winterOlympicsEventsCount, xlab = "Year", ylab = "Number of Events",

main = "Number of Events at Winter Olympics through years", type="l")

Screenshot of the line plot created in (6).

Chart, line chart

Description automatically generated

The entire R code used when creating the data frame in (7), scatter plot in (8), line in (9), and prediction in (10).

*# Data from 1984 onwards*

winterOlympicsSubset <- as.data.frame(winterOlympicsEventsCount[14:22])

names(winterOlympicsSubset)=c("Year", "Events")

winterOlympicsSubset$Year<-as.numeric(as.character(winterOlympicsSubset$Year))

*# Scatter plot of data from 1984 onwards*

plot(winterOlympicsSubset, xlab = "Year", ylab = "Number of Events",

main = "Number of Events at Winter Olympics from 1984 onwards")

*# Line fit*

linefit2 <- lm(Events~Year, data = winterOlympicsSubset)

abline(linefit2, col = "yellow", lwd = 4)

*# Predicting number of events in 2040*

predict(linefit2, list(Year=2040))

Screenshot of the scatter plot created in (8) with the line created in (9).Chart, line chart

Description automatically generated

The prediction (answer) made in (10).

A picture containing graphical user interface

Description automatically generated

The entire R code used when creating the scatter plot in (11), and line in (12).

marchTemps <- data.frame(temperature$YEAR, temperature$MAR)

names(marchTemps)[1:2] <- c("Year", "Temperature")

bloomDates <- data.frame(cherryBlossoms$Year,cherryBlossoms$Yoshino.peak.bloom.date)

names(bloomDates)[1:2] <- c("Year", "BloomDate")

mergedData <- merge(bloomDates, marchTemps,

by = intersect(names(marchTemps),

names(bloomDates)), by.x = 'Year')

*# Scatter plot*

plot(mergedData$Temperature, mergedData$BloomDate,

main = "Temperature on Peak Bloom Dates in March through years",

xlab = "Temperature on that day", ylab = "Peak Bloom Date",xlim = c(35, 60), ylim = c(70, 110), pch = 20)

*# Line plot*

linefit3 <- lm(mergedData$BloomDate ~ mergedData$Temperature)

abline(linefit3, col="red", lwd = 4)

Screenshot of the scatter plot created in (11) with the line created in (12).

Chart, scatter chart

Description automatically generated

**Your opinion about the correlation (or lack thereof) between the Cherry Blossom Peak Bloom Date and the Temperature in March.**

We can see that the line passes through the data points. We can say that the bloom date and temperature are correlated. As the temperature increases, the bloom date decreases. So higher temperature causes blooming to happen faster. But too high temperatures might not let blooming happen. So, there is a negative correlation between them.