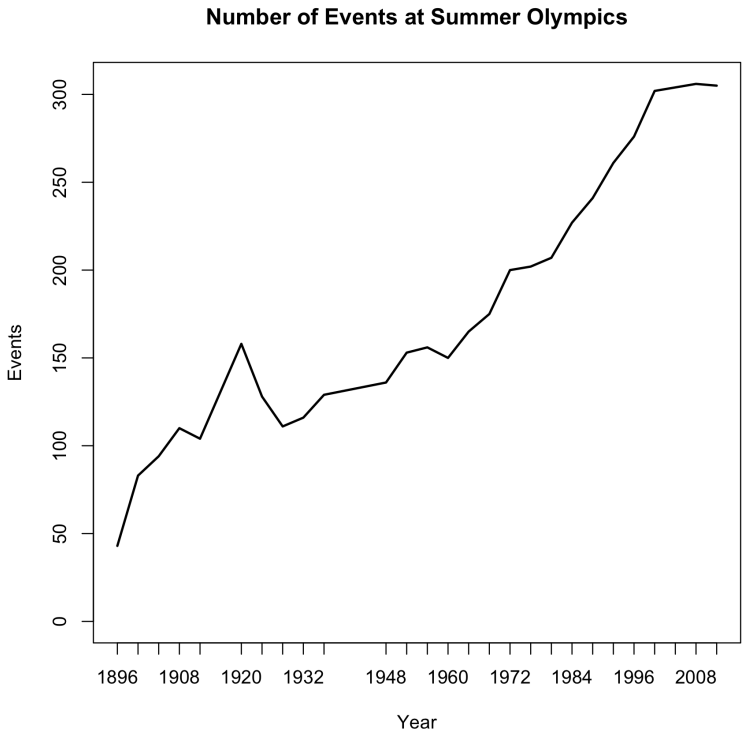
Name: Vikram Sahai Saxena
Net ID: vs799
RUID: 219004709

Module 9 Exercise: Line Plots and Linear Fits on Scatter Plots

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

```
a<-read.csv("summer.csv")
a$ev<-paste(a$Discipline, a$Event, a$Gender, sep="_")
b<-a[c("Year", "ev")]
c<-unique(b)
q<-table(c$Year)
plot(q, xlab="Year", ylab="Events", main="Number of Events at Summer Olympics",
type="l")
```

ii. **Screenshot of the line plot created in (1)**

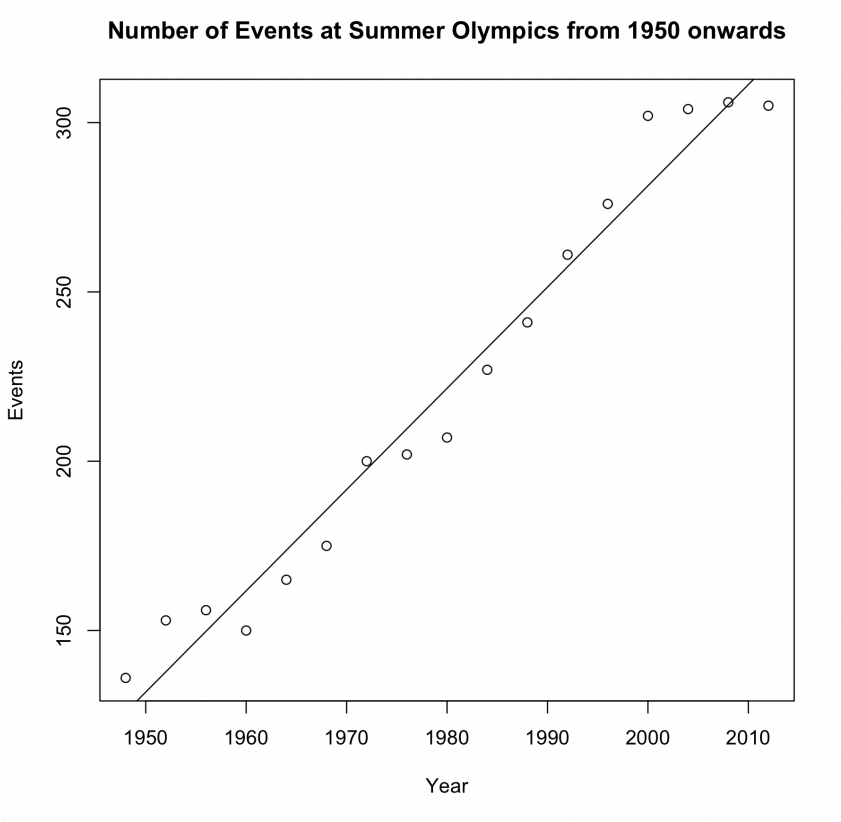


iii. **The entire R code used when creating the data frame in (2), scatter plot in (3), line in (4), and prediction in (5)**

```
qa<-as.data.frame(q[11:27])
names(qa)=c("Year", "Event")
qa$Year<-as.numeric(as.character(qa$Year))
plot(qa, xlab="Year", ylab="Events", main="Number of Events at Summer Olympics from
1950 onwards")
li_a<-lm(Event~Year, data=qa)
```

```
abline(li_a)
predict(li_a, list(Year=2040))
```

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



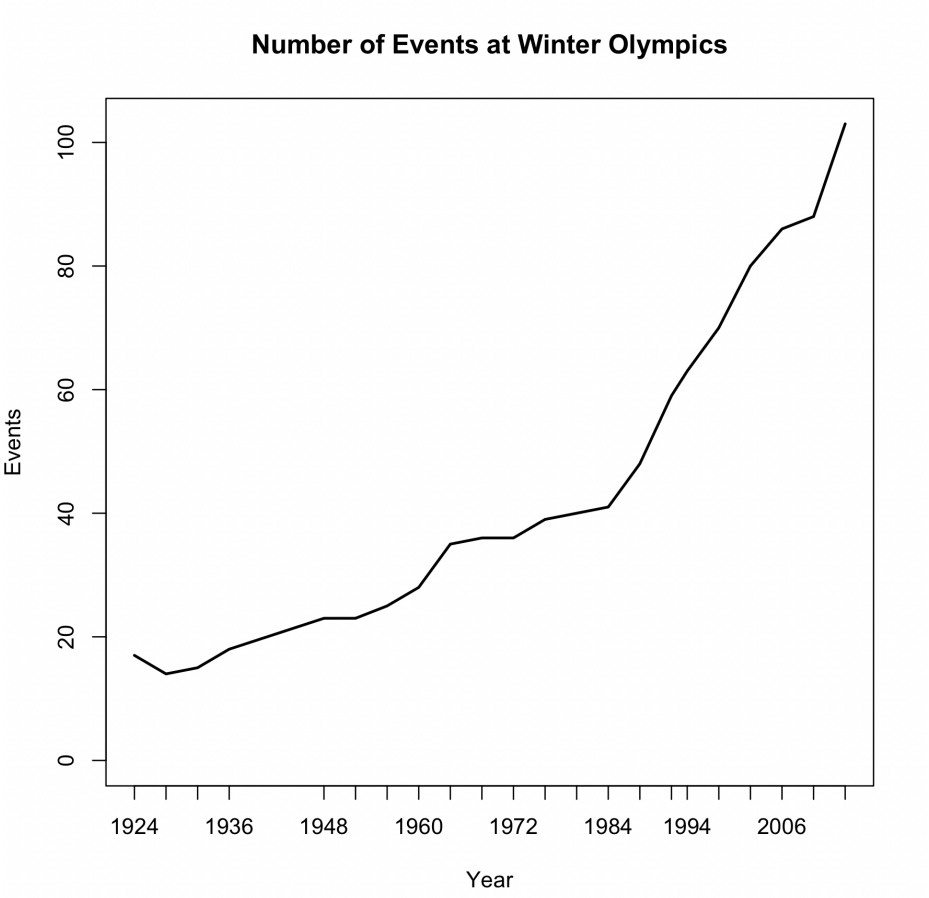
v. The prediction (answer) made in (5)

400.9412

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

```
d<-read.csv("winter.csv")
d$ev<-paste(d$Discipline, d$Event, d$Gender, sep="_")
e<-d[c("Year", "ev")]
f<-unique(e)
r<-table(f$Year)
plot(r, xlab="Year", ylab="Events", main="Number of Events at Winter Olympics",
type="l")
```

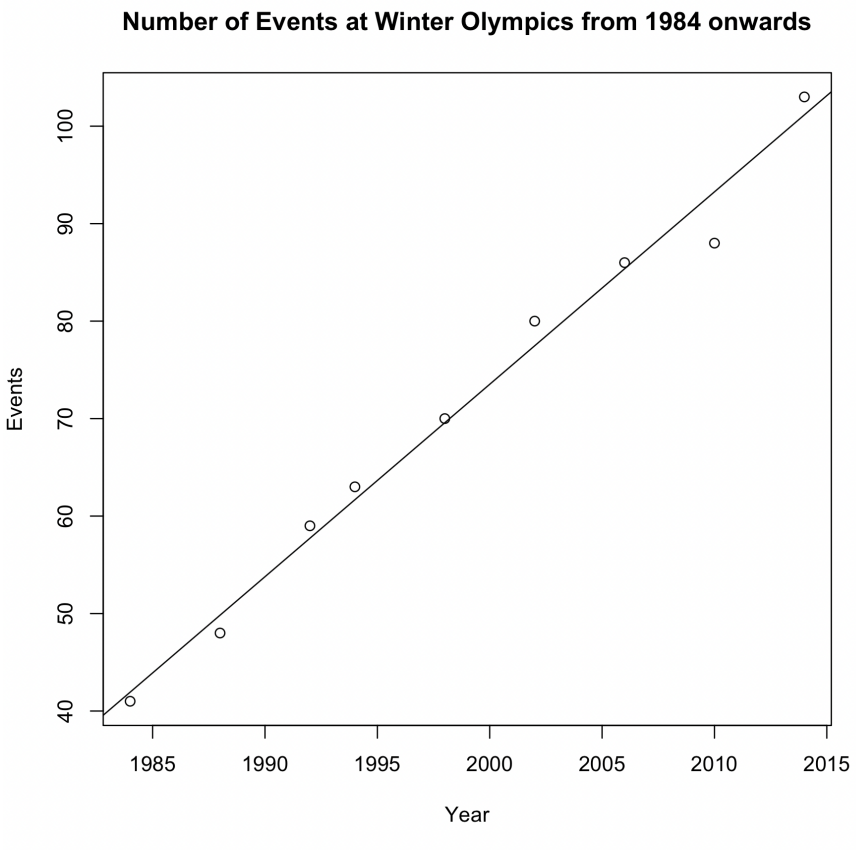
vii. Screenshot of the line plot created in (6)



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

```
ra<-as.data.frame(r[14:22])
names(ra)=c("Year", "Event")
ra$Year<-as.numeric(as.character(ra$Year))
plot(ra, xlab="Year", ylab="Events", main="Number of Events at Winter Olympics from 1984 onwards")
li_b<-lm(Event~Year, data=ra)
abline(li_b)
predict(li_b, list(Year=2040))
```

ix. Screenshot of the scatter plot created in (8) with the line created in (9)



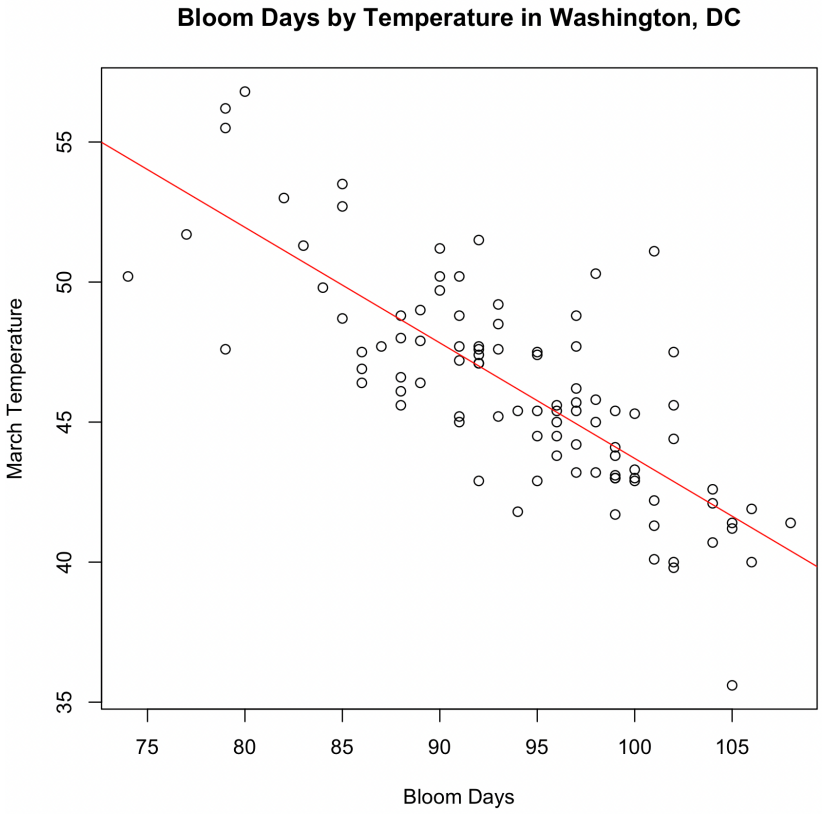
x. The prediction (answer) made in (10)

152.4854

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

```
cherry<-read.csv("cherry.csv")
washtemp<-read.csv("washtemp.csv", head=TRUE, sep='')
frame_cherry<-data.frame(cherry$Year, cherry$Yoshino.peak.bloom.date)
frame_washtemp<-data.frame(washtemp$YEAR, washtemp$MAR)
names(frame_cherry)[1:2]<-c("Year", "BloomDays")
names(frame_washtemp)[1:2]<-c("Year", "MarchTemp")
frameboth<-merge(frame_cherry, frame_washtemp, by =
intersect(names(frame_cherry), names(frame_washtemp)), by.x='Year')
plot(frameboth$BloomDays, frameboth$MarchTemp, xlab="Bloom Days", ylab="March
Temperature", main="Bloom Days by Temperature in Washington, DC")
abline(lm(frameboth$MarchTemp~frameboth$BloomDays), col="red")
```

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



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

The cherry blossom peak bloom date has a strong negative correlation with March temperatures, indicating earlier blooms in warmer March conditions and later ones in colder. However, correlation does not imply causation, and other variables may also affect bloom dates and temperatures, necessitating further analysis to clarify their relationship.