**DATA CLEANING**

To answer the first research question we need to plot a line graph (or histogram or scatterplot) and determine if the number of tornado occurrence has increased over the last 10 years. From the data we can see that the Midwest region has the seen the maximum number of tornadoes over the last 10 years, while the West region has seen the least.

//Read the .csv file

d=read.csv("C:\\INFM 600 Information Environments\\Project\\Data Sets \\Tornado Occurrences.csv")

//Create a subset of rows specific only to the midwest region

midwest = d[1:10,]

//Validate that the data in the subset is accurate

nrow(midwest)

head(midwest)

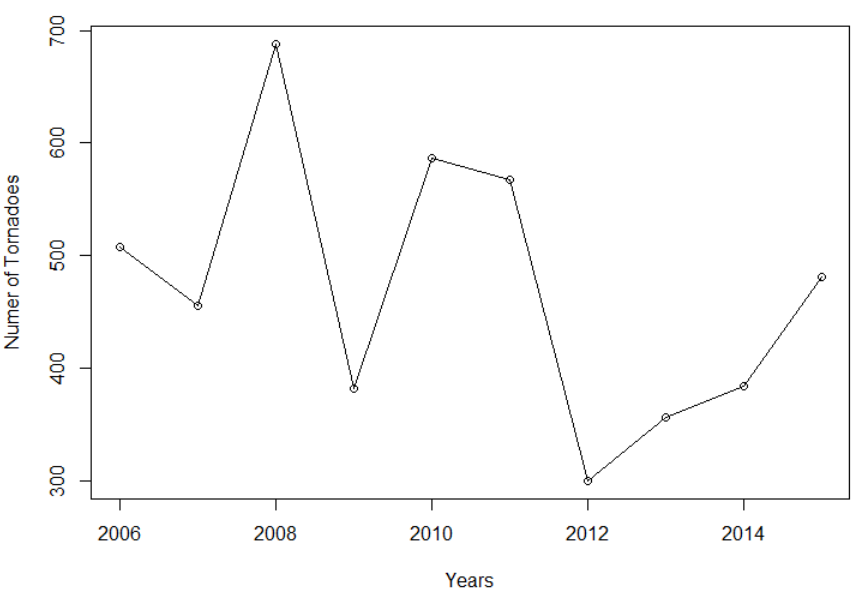
table(midwest)

//Get the mean of the number of tornado occurrences

mean(midwest$Total.Number.of.Tornado.Occurrences)

//Plot a line graph

plot(midwest$Year,midwest$Total.Number.of.Tornado.Occurrences,type="o")



This shows that there is no trend of tornado occurrences over the last 10 years in the midwest region.

Now we will look at similar data in the West region.

//Create a subset of data for the West region

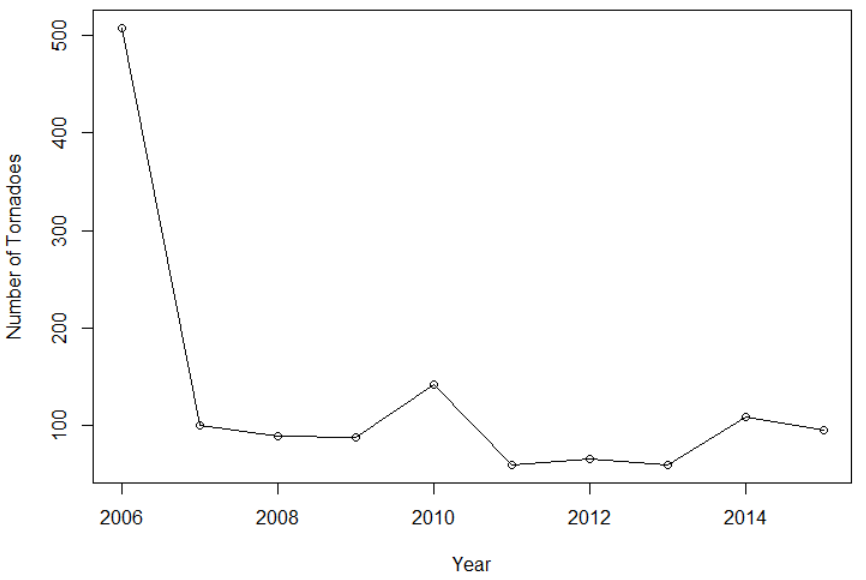
west = d[c(1,32,33,34,35,36,37,38,39,40),]

//Calculate the mean

mean(west$Total.Number.of.Tornado.Occurrences,na.rm=TRUE)

//Plot the line graph

plot(west$Year,west$Total.Number.of.Tornado.Occurrences,type="o")



Again, we can see that there is no trend of tornado occurrences even in the West region.