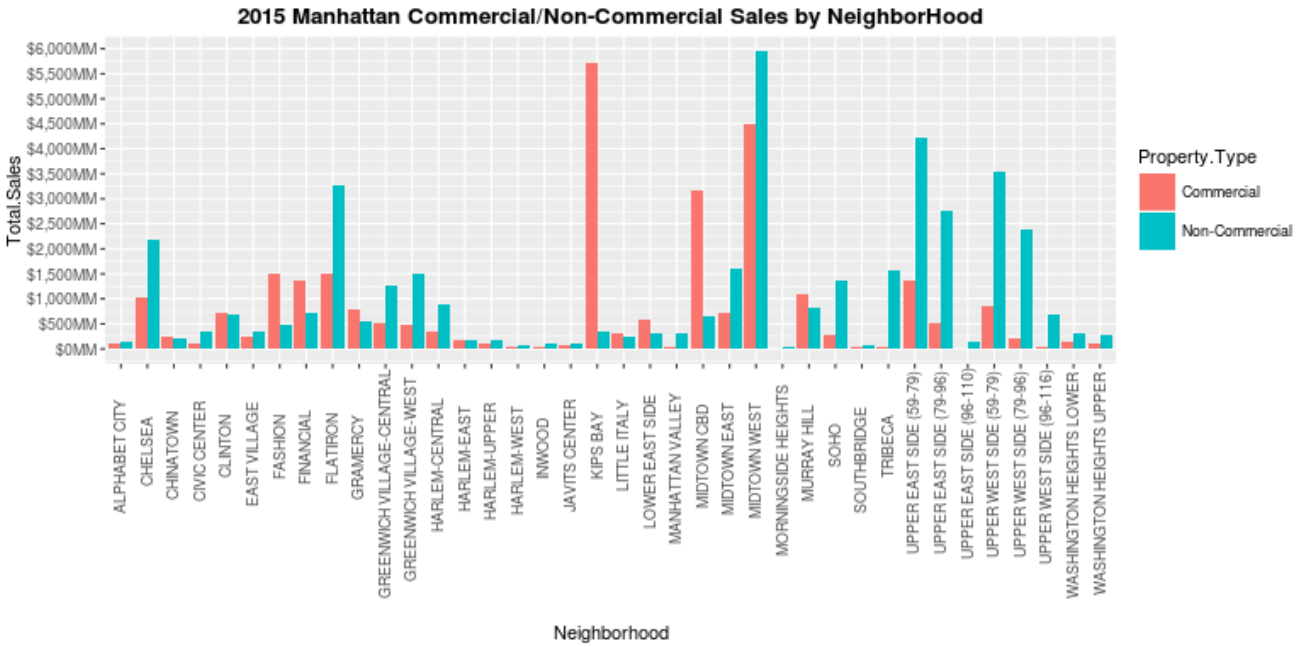
ROLLING-SALES MANHATTAN 2015

```
## This is an exercise on processing data and creating meaningful visuals using R. Here I have a
csv file that contains data about 2015 property sales in Manhattan.

## First we load the csv into a data frame and create a list of each column header and its class.

DF <- read.csv('rollingsales_manhattan.csv')
myList <-list()
for(i in 1:ncol(DF)){
myList[[colnames(DF)[i]]]<-class(DF[[i]])
}
myList

## Now we can cast the variables to appropriate classes.

DF$ADDRESS<-as.character(DF$ADDRESS)
DF$APARTMENT.NUMBER<-as.character(DF$APARTMENT.NUMBER)
DF$COMMERCIAL.UNITS <- gsub(",","",DF$COMMERCIAL.UNITS) ## Remove commas before changing
DF$COMMERCIAL.UNITS <- as.numeric(DF$COMMERCIAL.UNITS)     to a numeric class.
DF$TOTAL.UNITS <- gsub(",","",DF$TOTAL.UNITS)
DF$TOTAL.UNITS<- as.numeric(DF$TOTAL.UNITS)
DF$LAND.SQUARE.FEET <- gsub(",","",DF$LAND.SQUARE.FEET)
DF$LAND.SQUARE.FEET<- as.numeric(DF$LAND.SQUARE.FEET)
DF$GROSS.SQUARE.FEET <- gsub(",","",DF$GROSS.SQUARE.FEET)
DF$GROSS.SQUARE.FEET<- as.numeric(DF$GROSS.SQUARE.FEET)
DF$SALE.PRICE<- gsub(",","",DF$SALE.PRICE)
DF$SALE.PRICE<- gsub("\\$","",DF$SALE.PRICE) ## We need to remove the $ as well.
DF$SALE.PRICE <- as.numeric(DF$SALE.PRICE)
DF$SALE.DATE<-as.character.Date(DF$SALE.DATE)

## A $0 Sales.Price indicates that there was a transfer of ownership without a cash consideration,
so we will remove all rows from the data with a Sales.Price of $0.

DF <- subset(DF, SALE.PRICE != 0)

## Now split the data into Commercial and Non-Commercial properties and aggregate the total sales
for each split. We will consider any property with at least one Commercial unit to be a Commercial
property regardless of Residential units.

Commercial <- subset(DF, COMMERCIAL.UNITS > 0)
NonCommercial <- subset(DF, COMMERCIAL.UNITS == 0)
CommercialNeighborhood<-aggregate(Commercial$SALE.PRICE,list(Commercial$NEIGHBORHOOD),sum)
NonCommercialNeighborhood<-aggregate(NonCommercial$SALE.PRICE,list(NonCommercial$NEIGHBORHOOD),sum)
colnames(NonCommercialNeighborhood)[c(1,2)]<-c("Neighborhood","Total.Sales") ## Keep headers
colnames(CommercialNeighborhood)[c(1,2)]<-c("Neighborhood","Total.Sales")       the same.

## We will now create a graph to display the total sales of Commercial and Non-Commercial
properties in each Neighborhood of Manhattan. First mark each of our aggregated splits as either
Commercial or Non-Commercial with a new column, Property.Type.

CommercialNeighborhood$Property.Type<-rep('Commercial',nrow(CommercialNeighborhood))
NonCommercialNeighborhood$Property.Type<-rep('Non-Commercial',nrow(NonCommercialNeighborhood))

## Then combine together and change Property.Type to a factor.

NBSalesByPropType<-rbind(CommercialNeighborhood,NonCommercialNeighborhood)
NBSalesByPropType$Property.Type<-as.factor(NBSalesByPropType$Property.Type)

## We now create a graph using ggplot2, as you can see there are many formatting options provided
to allow you to achieve great levels of customization.

library(“ggplot2”)  ## You need ggplot2 installed.
library("scales")  ## For formating.
TSLabels<-paste("$",format(seq(0,6000,500),big.mark=",",trim=TRUE),"MM",sep="") ## y-axis label.
ggplot(NBSalesByPropType, aes(x=Neighborhood,y=Total.Sales, fill=Property.Type)) +
```

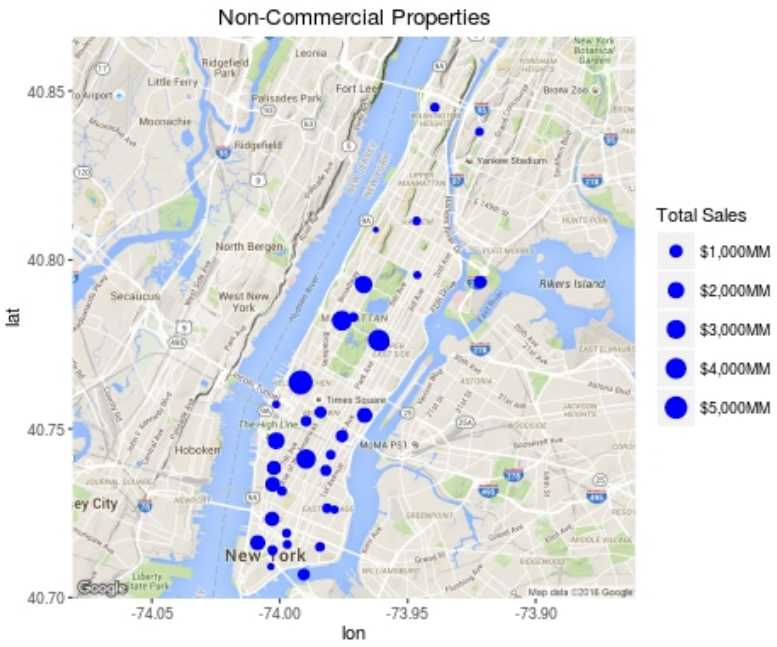
```
geom_bar(stat='identity',position = position_dodge()) + theme(axis.text.x = element_text(angle=90))
+theme(axis.text.x=element_text(hjust=1.15,vjust=.4))+scale_y_continuous(labels =
TSLabels,breaks=seq(0,6000000000,500000000))+ ggtitle("2015 Manhattan Commercial/Non-Commercial
Sales by NeighborHood") + theme(plot.title = element_text(lineheight=.8, face="bold"))
```

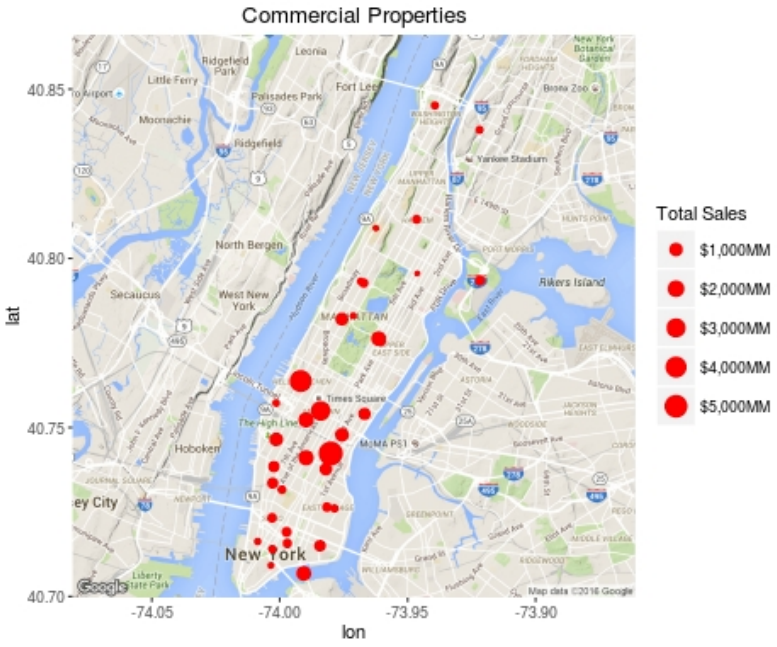


From this graph we can see which Neighborhoods are home to the most Commercial and Non-Commercial property sales activity in 2015. Kips Bay had the highest total sales for Commercial properties by a large margin with over 1 Billion more in sales than the next highest neighborhood. Midtown West had the highest Non-Commercial total sales and the most total sales overall. This visual is a very broad outlook on one aspect of our data, but it may provide direction in moving towards more detailed investigation. For instance we may want to see what other attributes may be correlated with total sales by Neighborhood such as location, the building class category, or even the sale date, maybe some financial event effected played a role?

Lets look at a geographical distribution of total sales of both Commercial and Non-Commercial properties in Manhattan. We will use ggmap and the size of the dot at the location of each Neighborhood will be correlated with the total sales amount.

```
library(ggmap)
NonComNames<-paste(NonCommercialNeighborhood$Neighborhood,", manhattan,NY") ## Makes ggmap's search
ComNames<-paste(CommercialNeighborhood$Neighborhood,", manhattan,NY")          more accurate.
NonComNBgeocode<-geocode(NonComNames,output="latlon")  ## Coordinates for each Neighborhood.
ComNBgeocode<-geocode(ComNames,output="latlon")
NonComNBgeocode$Total.Sales<-NonCommercialNeighborhood$Total.Sales
ComNBgeocode$Total.Sales<-CommercialNeighborhood$Total.Sales
map=get_map(location = 'MANHATTAN,NY', zoom = 12)
SizeLabels<-paste("$",format(seq(1000,5000,1000),big.mark=",",trim=TRUE),"MM",sep="")
NonComMap<-ggmap(map)
+geom_point(data=NonComNBgeocode,aes(x=lon,y=lat,size=Total.Sales),colour="blue")+ggtitle("Non-
Commercial Properties")+scale_size_continuous(name="Total
Sales",breaks=seq(1000000000,5000000000,1000000000),labels=SizeLabels)
ComMap<-ggmap(map)+geom_point(data=ComNBgeocode,aes(x=lon,y=lat,size=Total.Sales),colour="red")
+ggtitle("Commercial Properties")+scale_size_continuous(name="Total
Sales",breaks=seq(1000000000,5000000000,1000000000),labels=SizeLabels)
NonComMap
ComMap
```


Non-Commercial Properties
lat
lon
40.85
40.80
40.75
40.70
-74.05
-74.00
-73.95
-73.90
New York
Hoboken
Secaucus
North Bergen
Ridgefield
Fort Lee
Rikers Island
Total Sales
$1,000MM
$2,000MM
$3,000MM
$4,000MM
$5,000MM


Commercial Properties
lat
lon
40.85
40.80
40.75
40.70
-74.05
-74.00
-73.95
-73.90
New York
Hoboken
Secaucus
North Bergen
Ridgefield
Fort Lee
Rikers Island
Total Sales
$1,000MM
$2,000MM
$3,000MM
$4,000MM
$5,000MM