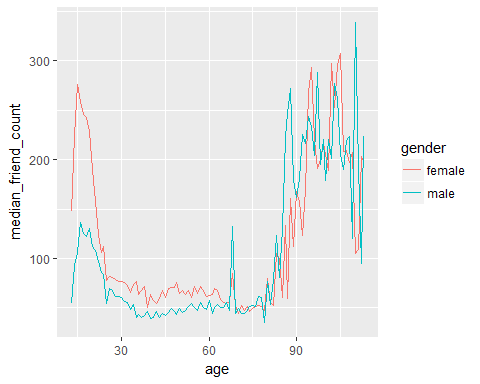
Facebook3Var

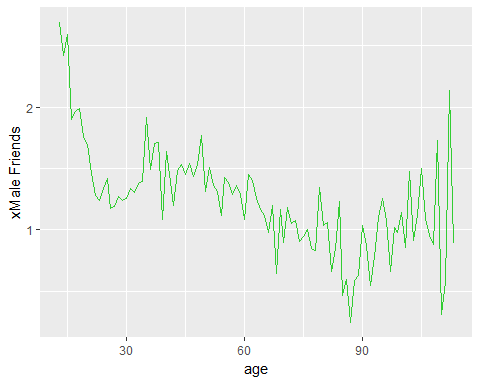
#Grouping by age and gender  
pf.fc\_by\_age\_gender <- pf %>%  
 filter(!is.na(gender)) %>%  
 group\_by(age, gender) %>%  
 summarise(mean\_friend\_count = mean(friend\_count),  
 median\_friend\_count = median(friend\_count),  
 n = n()) %>%  
 ungroup() %>%  
 arrange(age)

#putting color in an aesthetic wrapper in the line  
ggplot(pf.fc\_by\_age\_gender, aes(x=age, y=median\_friend\_count)) +geom\_line(aes(color=gender))



#Creating a wide dataframe  
library(reshape2)  
pf.fc\_by\_age\_gender.wide <- dcast(pf.fc\_by\_age\_gender,  
 age ~ gender,  
 value.var = 'median\_friend\_count')

#Plotting the multiple of friends that females have compared to men  
ggplot(pf.fc\_by\_age\_gender.wide, aes(x=age, y=female/male)) +geom\_line(color='lime green')+  
 ylab('xMale Friends')



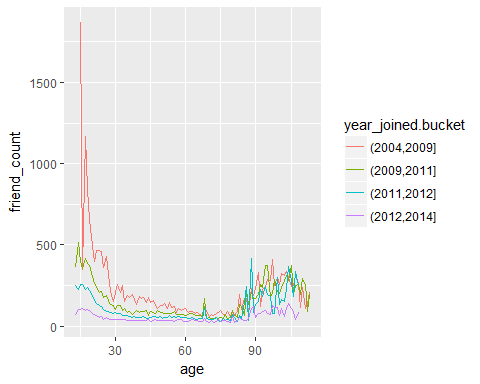
#Creating a variable year joined by subtracting tenure(days) from date of sample (2014)   
pf$year\_joined <- floor(2014 - (pf$tenure/365))   
summary(pf$year\_joined)

## Min. 1st Qu. Median Mean 3rd Qu. Max. NA's   
## 2005 2012 2012 2012 2013 2014 2

pf$year\_joined.bucket <- cut(pf$year\_joined, breaks = c(2004,2009,2011,2012, 2014))   
table(pf$year\_joined.bucket)

##   
## (2004,2009] (2009,2011] (2011,2012] (2012,2014]   
## 6669 15308 33366 43658

#Plotting the different lines for each year joined bucket   
ggplot(subset(pf, !is.na(pf$year\_joined.bucket)), aes(x=age, y=friend\_count))+  
 geom\_line(aes(color=year\_joined.bucket), stat='summary', fun.y=median)



#Plotting the mean but also adding the grandmean   
ggplot(subset(pf, !is.na(pf$year\_joined.bucket)), aes(x=age, y=friend\_count))+   
 geom\_line(aes(color=year\_joined.bucket), stat='summary', fun.y=mean)+   
 geom\_line(stat='summary', fun.y=mean)

