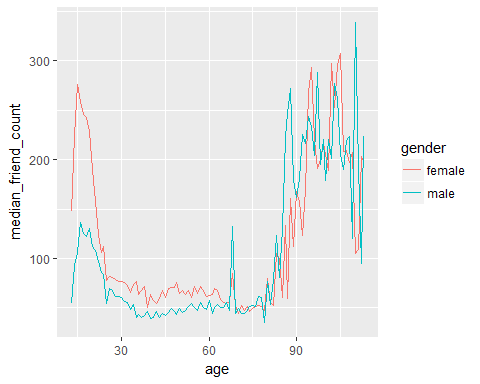
Facebook3Var

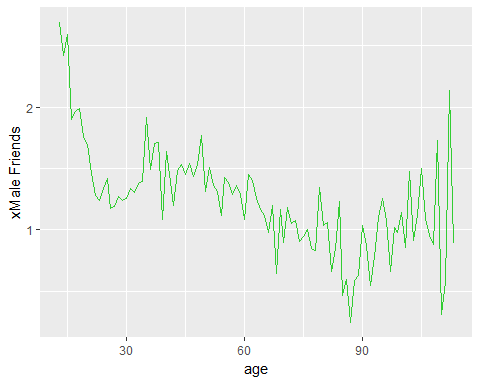
#Grouping by age and gender  
pf.fc\_by\_age\_gender <- pf %>%  
 filter(!is.na(gender)) %>%  
 group\_by(age, gender) %>%  
 summarise(mean\_friend\_count = mean(friend\_count),  
 median\_friend\_count = median(friend\_count),  
 n = n()) %>%  
 ungroup() %>%  
 arrange(age)

#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')



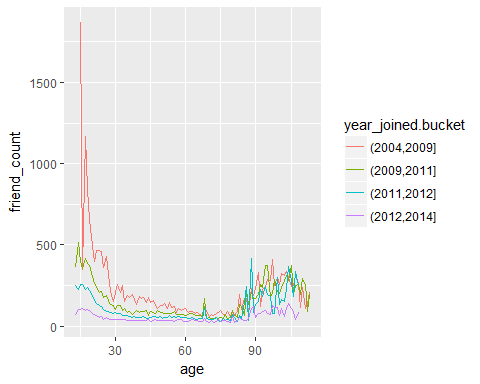
#Creating a variable year joined by subtracting tenure(days) from date of sample (2014)   
pf$year\_joined <- floor(2014 - (pf$tenure/365))   
summary(pf$year\_joined)

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

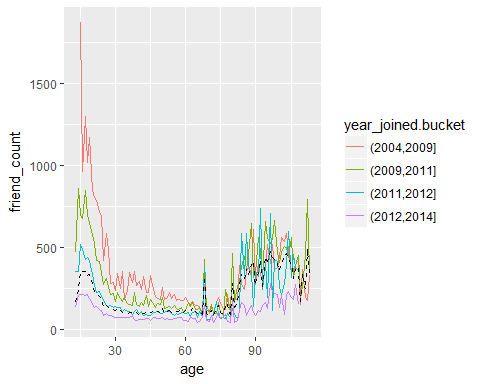
pf$year\_joined.bucket <- cut(pf$year\_joined, breaks = c(2004,2009,2011,2012, 2014))   
table(pf$year\_joined.bucket)

##   
## (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, linetype=2)

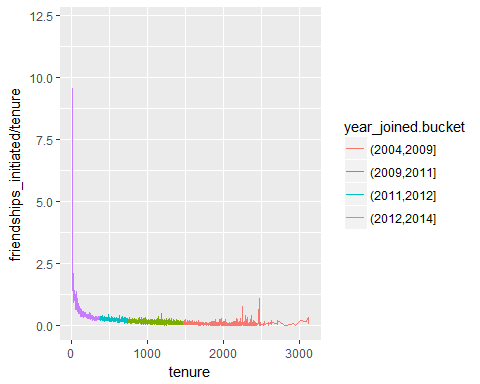


#If we were to look at the fc by day as a rate: (fc/tenure)  
with(subset(pf, tenure>0), summary(friend\_count/tenure))

## Min. 1st Qu. Median Mean 3rd Qu. Max.   
## 0.0000 0.0775 0.2205 0.6096 0.5658 417.0000

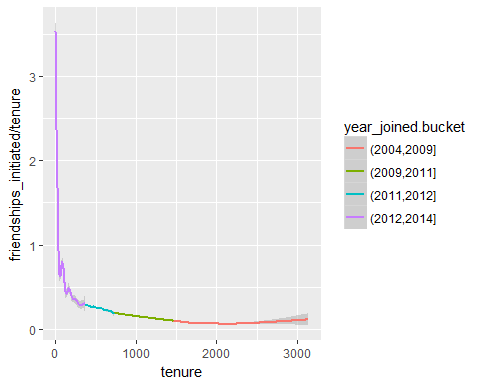
#looking at how many friendships someone initiates based on tenure, colored by year joined  
ggplot(subset(pf, tenure>0), aes(x=tenure, y=friendships\_initiated/tenure)) +  
 geom\_line(aes(color=year\_joined.bucket), stat='summary')

## No summary function supplied, defaulting to `mean\_se()



#With smooth line (using loess)  
ggplot(subset(pf, tenure>0), aes(x=tenure, y=friendships\_initiated/tenure)) +  
 geom\_smooth(aes(color=year\_joined.bucket))

## `geom\_smooth()` using method = 'gam'



library(GGally)

##   
## Attaching package: 'GGally'

## The following object is masked from 'package:dplyr':  
##   
## nasa

theme\_set(theme\_minimal(20))

#Using a random seed and generating a scatterplot matrix  
set.seed(1836)  
pf\_subset <- pf[, c(2:15)]  
names(pf\_subset)

## [1] "age" "dob\_day"   
## [3] "dob\_year" "dob\_month"   
## [5] "gender" "tenure"   
## [7] "friend\_count" "friendships\_initiated"  
## [9] "likes" "likes\_received"   
## [11] "mobile\_likes" "mobile\_likes\_received"  
## [13] "www\_likes" "www\_likes\_received"

ggpairs(pf\_subset[sample.int(nrow(pf\_subset), 1000),])

## `stat\_bin()` using `bins = 30`. Pick better value with `binwidth`.  
## `stat\_bin()` using `bins = 30`. Pick better value with `binwidth`.  
## `stat\_bin()` using `bins = 30`. Pick better value with `binwidth`.  
## `stat\_bin()` using `bins = 30`. Pick better value with `binwidth`.

## Warning: Removed 2 rows containing non-finite values (stat\_boxplot).  
  
## Warning: Removed 2 rows containing non-finite values (stat\_boxplot).  
  
## Warning: Removed 2 rows containing non-finite values (stat\_boxplot).  
  
## Warning: Removed 2 rows containing non-finite values (stat\_boxplot).  
  
## Warning: Removed 2 rows containing non-finite values (stat\_boxplot).  
  
## Warning: Removed 2 rows containing non-finite values (stat\_boxplot).  
  
## Warning: Removed 2 rows containing non-finite values (stat\_boxplot).  
  
## Warning: Removed 2 rows containing non-finite values (stat\_boxplot).  
  
## Warning: Removed 2 rows containing non-finite values (stat\_boxplot).

## `stat\_bin()` using `bins = 30`. Pick better value with `binwidth`.  
## `stat\_bin()` using `bins = 30`. Pick better value with `binwidth`.  
## `stat\_bin()` using `bins = 30`. Pick better value with `binwidth`.  
## `stat\_bin()` using `bins = 30`. Pick better value with `binwidth`.  
## `stat\_bin()` using `bins = 30`. Pick better value with `binwidth`.  
## `stat\_bin()` using `bins = 30`. Pick better value with `binwidth`.  
## `stat\_bin()` using `bins = 30`. Pick better value with `binwidth`.  
## `stat\_bin()` using `bins = 30`. Pick better value with `binwidth`.  
## `stat\_bin()` using `bins = 30`. Pick better value with `binwidth`.

