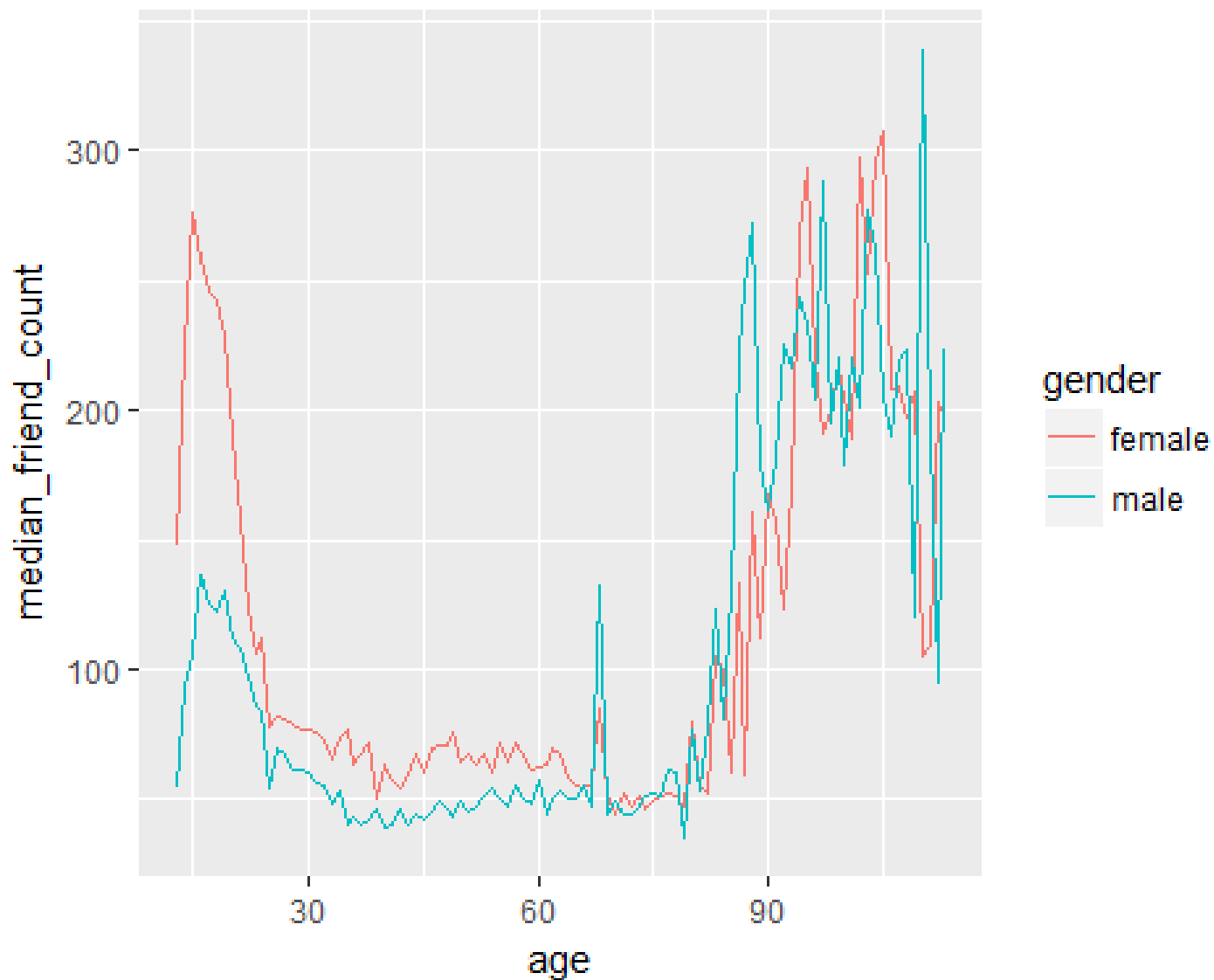


Facebook

```
#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))
```



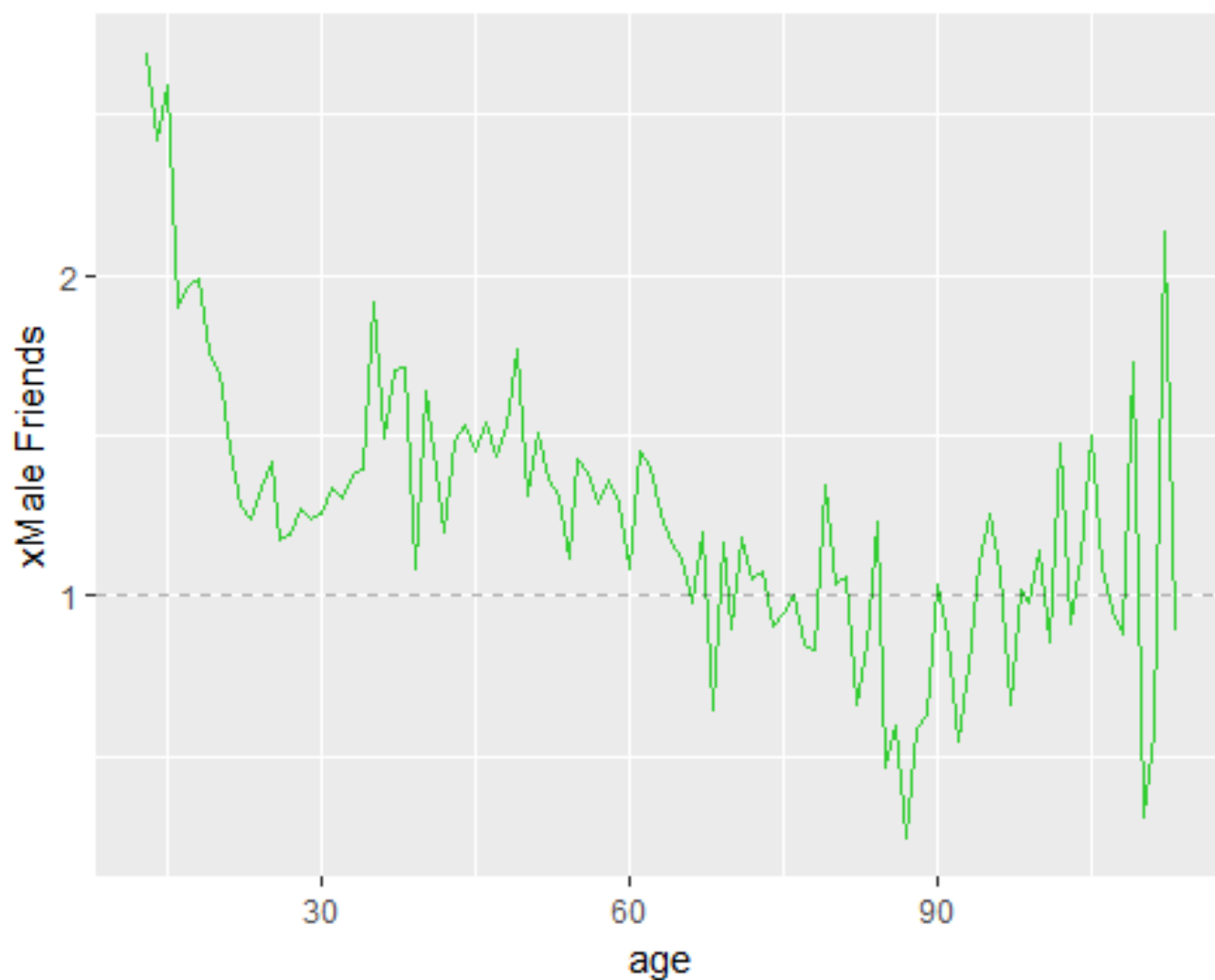
#Creating a wide dataframe

```
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') +  
  geom_hline(yintercept = 1, alpha =0.3, linetype=2)+  
  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
```

```
table(pf$year_joined)
```

```
## 2005 2006 2007 2008 2009 2010 2011 2012 2013 2014
##    9   15  581 1507 4557 5448 9860 33366 43588   70
```

```
#Dividing up the years joined with cut function
```

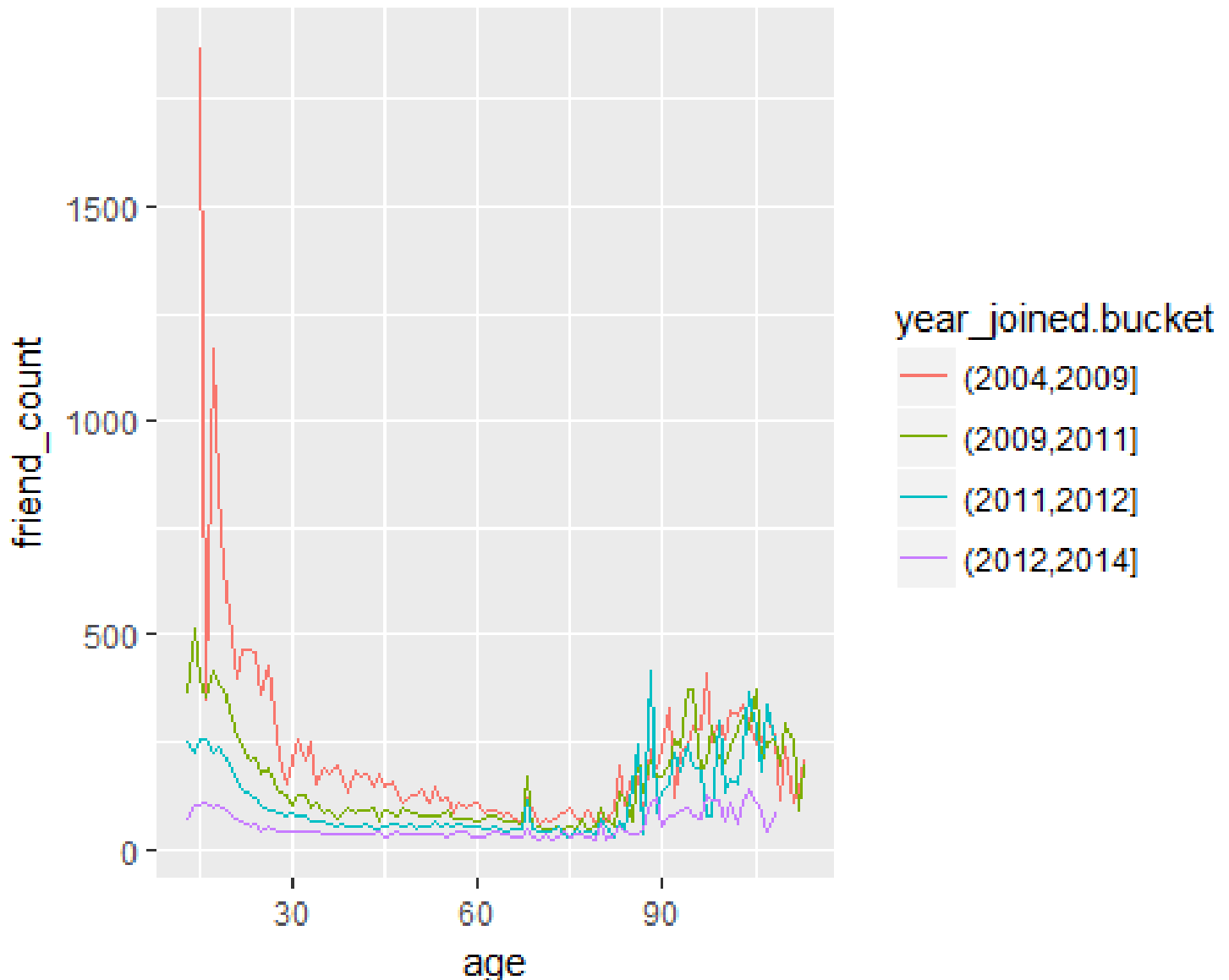
```
pf$year_joined.bucket <- cut(pf$year_joined, breaks = c(2004,2009,2011,2012, 2014))
```

```
table(pf$year_joined.bucket)
```

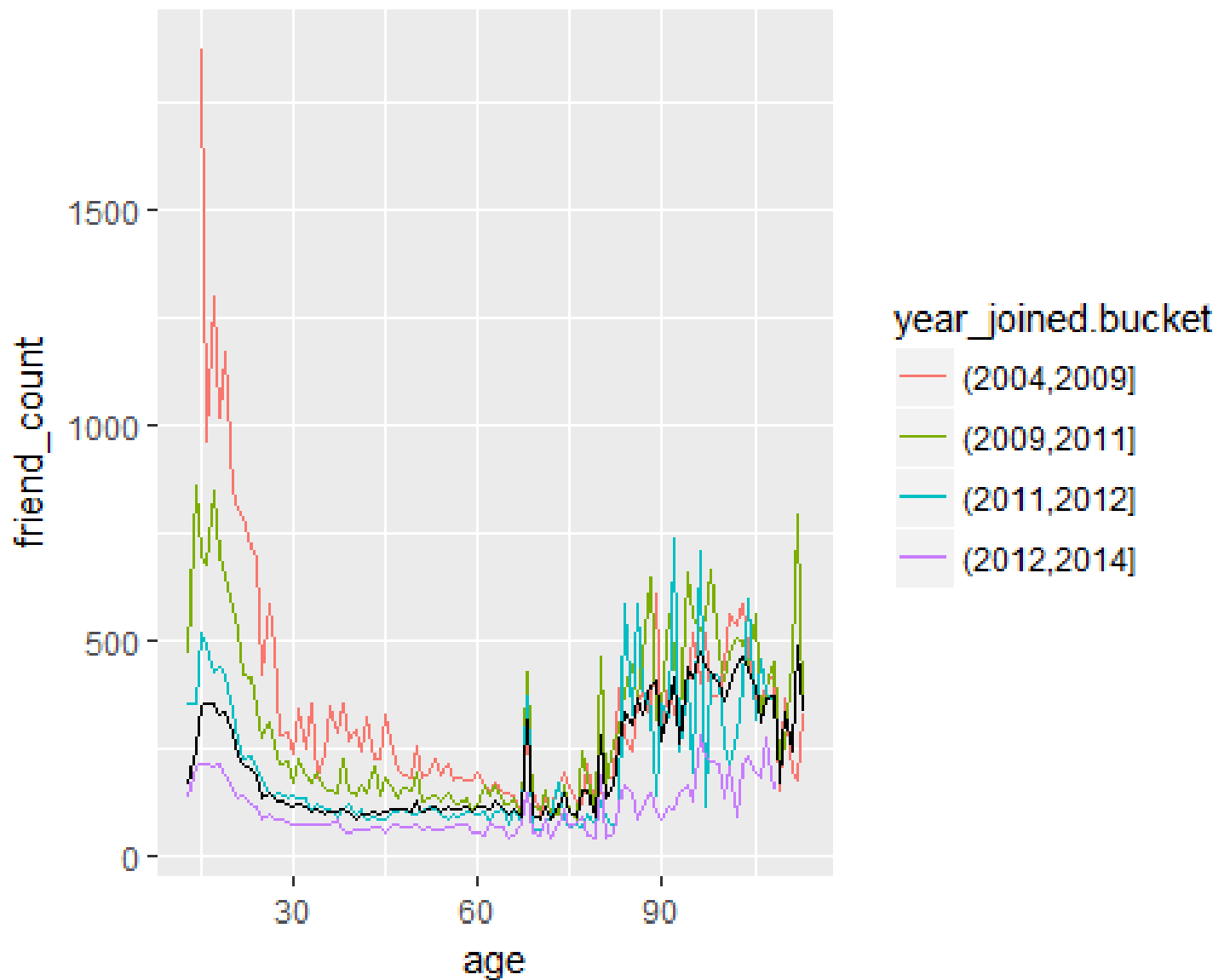
```
## (2004,2009] (2009,2011] (2011,2012] (2012,2014]
##      6669      15308      33366      43658
```

```
#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)
```

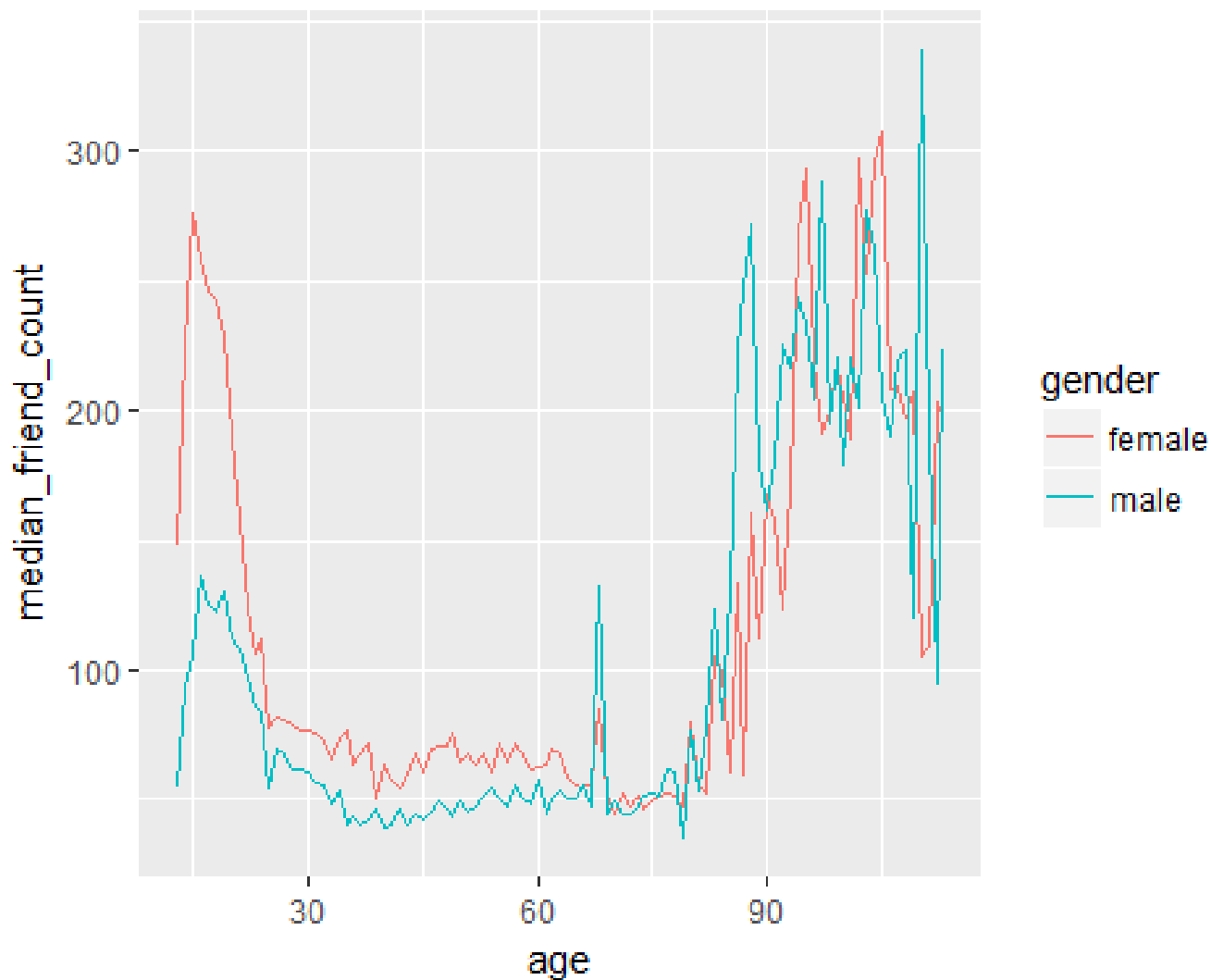


```

#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))

```

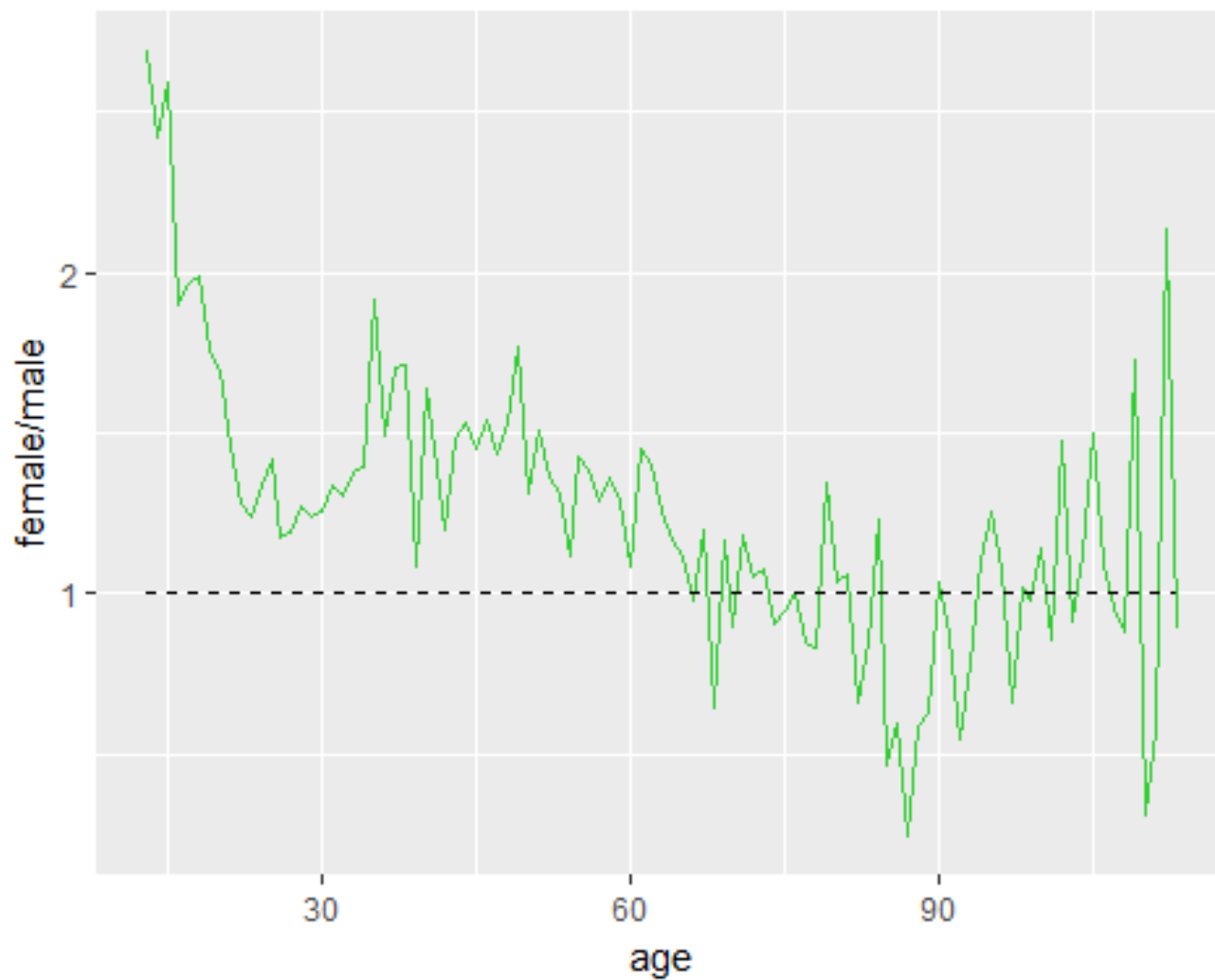


```

#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')+
  geom_line(y=1, linetype=2)

```



```

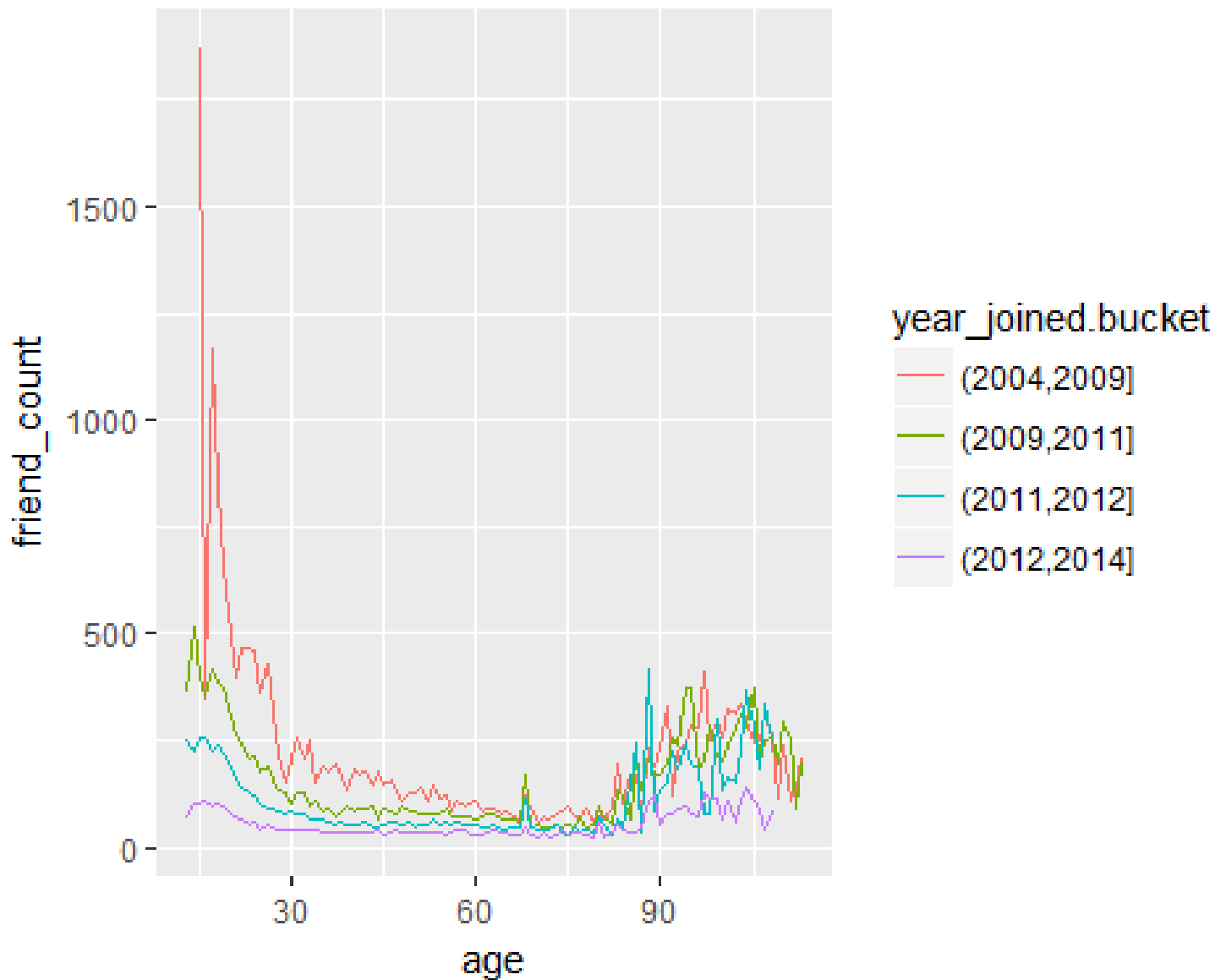
#Creating a variable year joined by subtracting tenure(days) from date of sample (2014)
pf$year_joined <- floor(2014 - (pf$tenure/365))
summary(pf$year_joined)

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

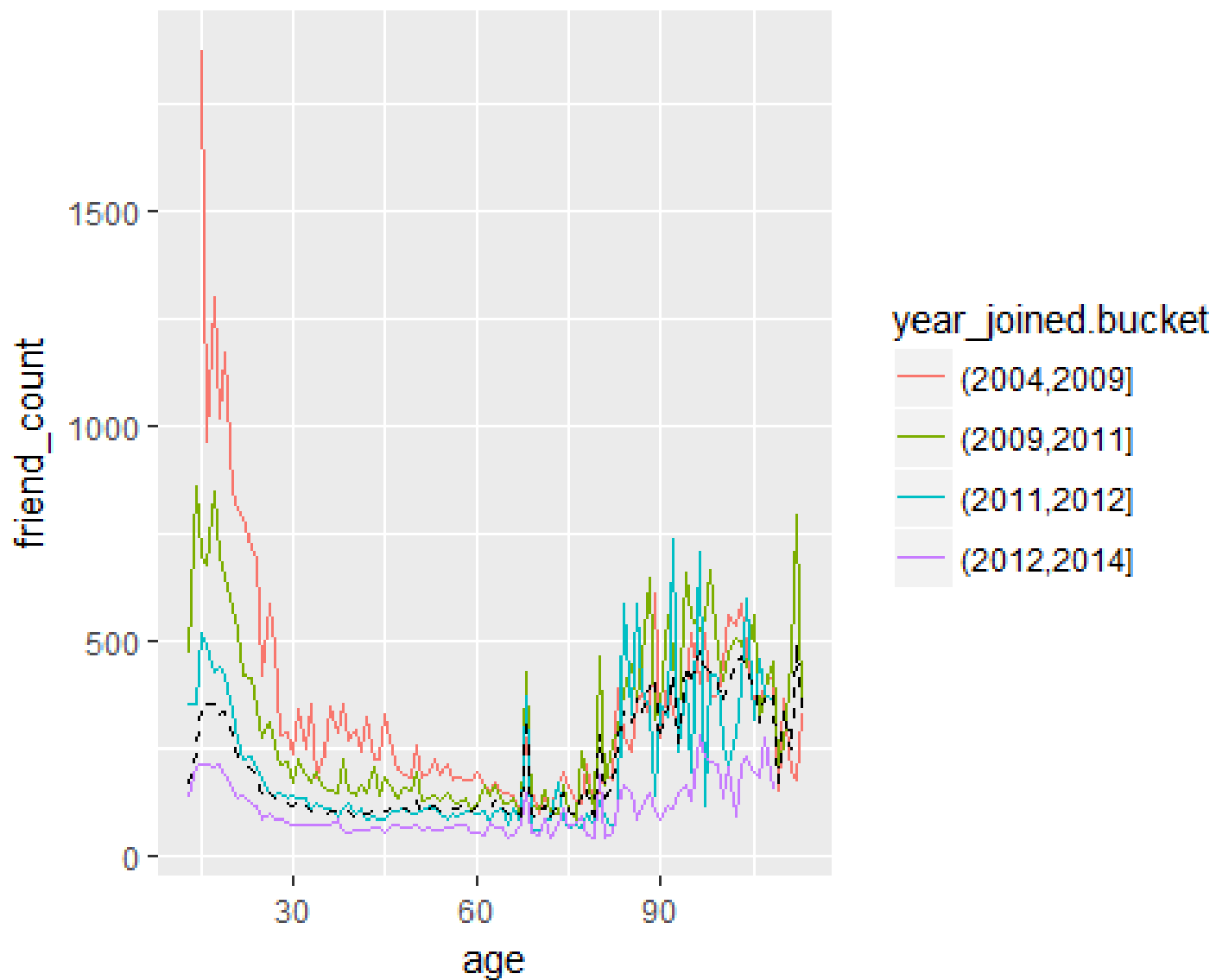
pf$year_joined.bucket <- cut(pf$year_joined, breaks = c(2004,2009,2011,2012, 2014))
table(pf$year_joined.bucket)
## (2004,2009] (2009,2011] (2011,2012] (2012,2014]
##      6669      15308      33366      43658

#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)
```




```

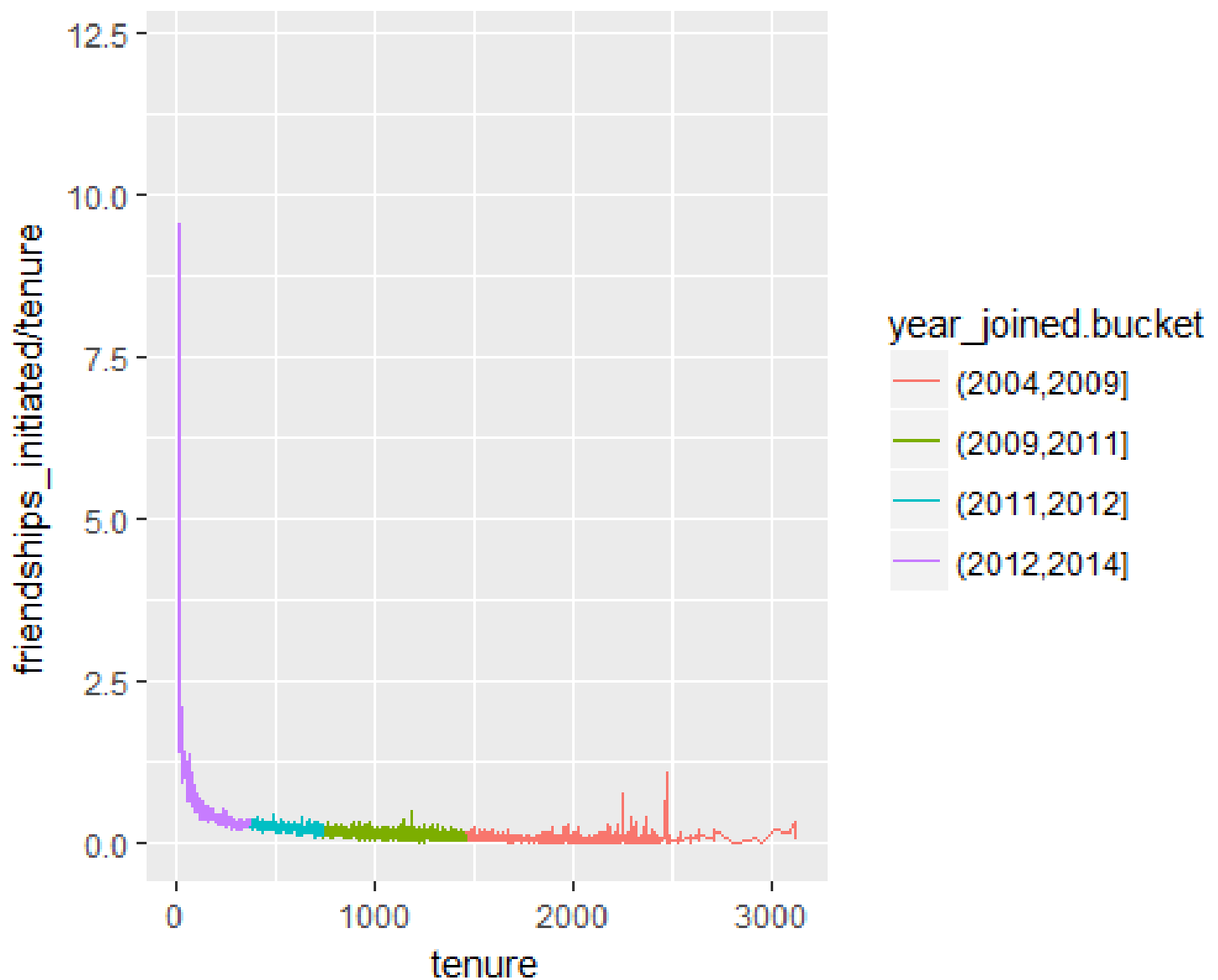
#If we were to look at the fc by day as a rate: (fc/tenure)
with(subset(pf, tenure>0), summary(friend_count/tenure))

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

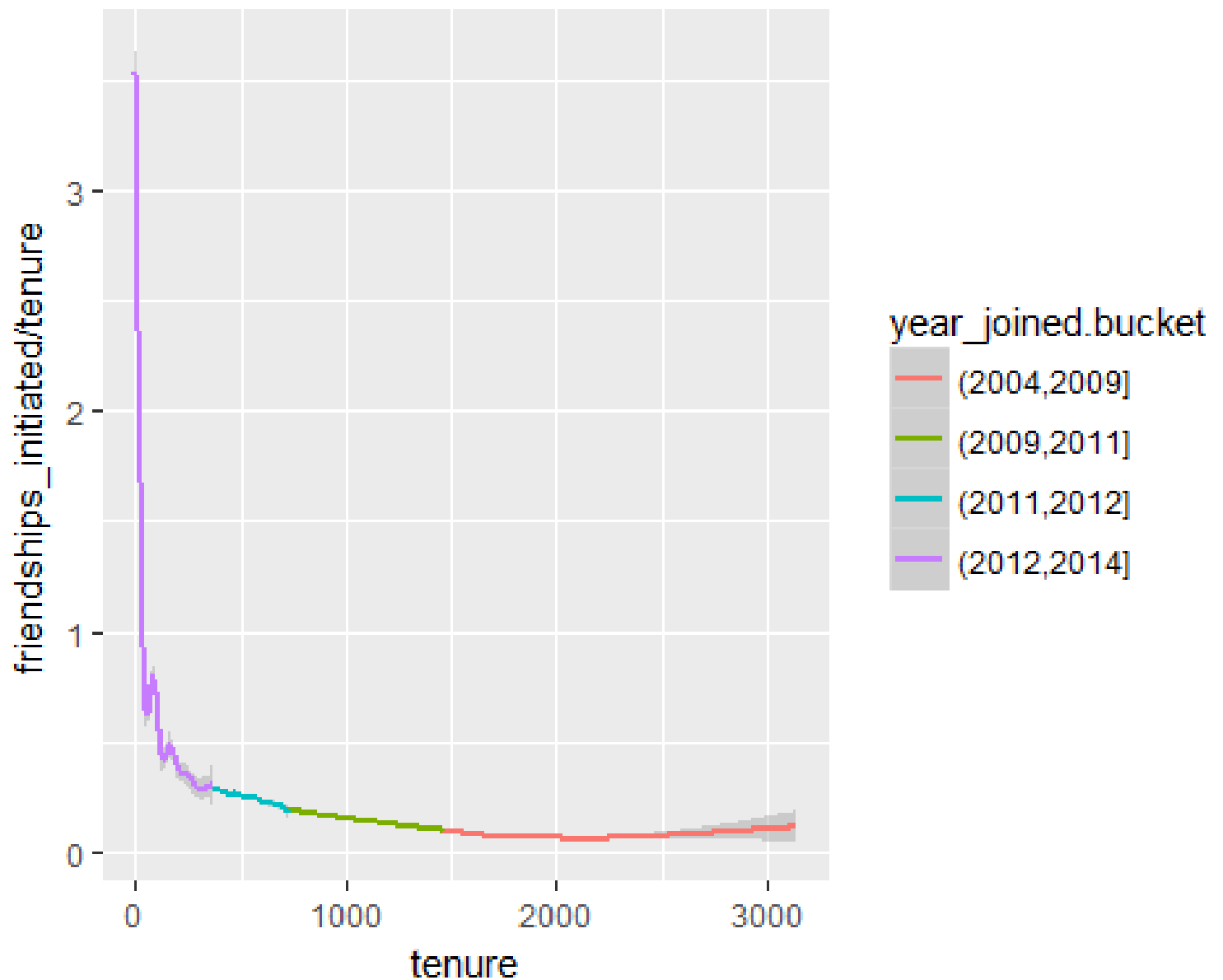
#Looking at how many friendships someone initiates based on tenure, colored by year
joined
ggplot(subset(pf, tenure>0), aes(x=tenure, y=friendships_initiated/tenure)) +
  geom_line(aes(color=year_joined.bucket), stat='summary')

## 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'
```



```

pf <- transform(pf, prop_initiated = ifelse(friend_count>0,
friendships_initiated/friend_count, 0))

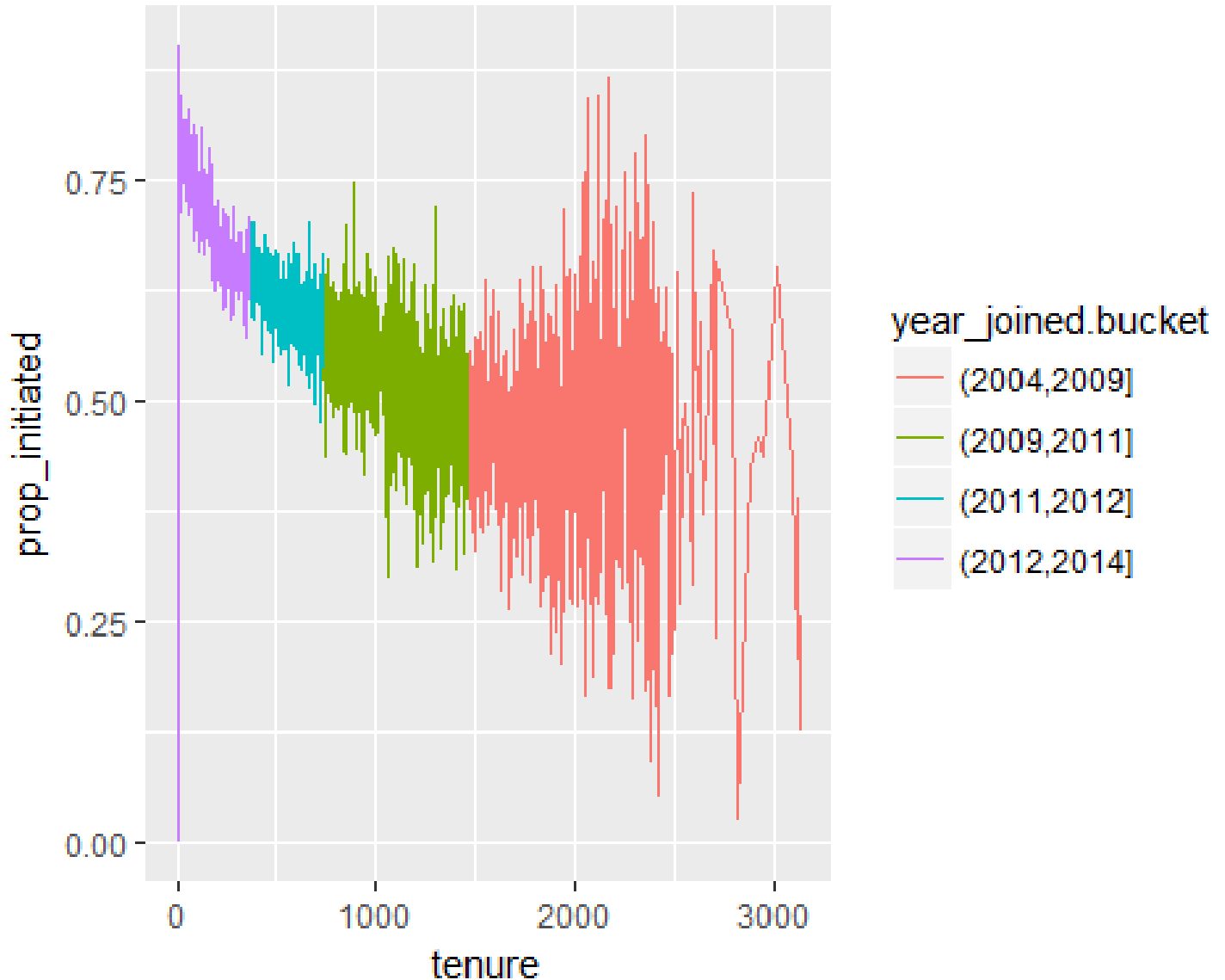
summary(pf$prop_initiated)

##      Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
## 0.0000  0.4400  0.6184  0.5958  0.7795  1.0000

#plotting the proportion of friendships initiated over tenure
ggplot(pf, aes(x=tenure, y=prop_initiated)) +
  geom_line(aes(color=year_joined.bucket), stat='summary', fun.y=median)

## Warning: Removed 2 rows containing non-finite values (stat_summary).

```



#Looking at it wiht a smooth line

```
ggplot(pf, aes(x=tenure, y=prop_initiated)) +  
  geom_smooth(aes(color=year_joined.bucket))
```

```
## `geom_smooth()` using method = 'gam'
```

```
## Warning: Removed 2 rows containing non-finite values (stat_smooth).
```

