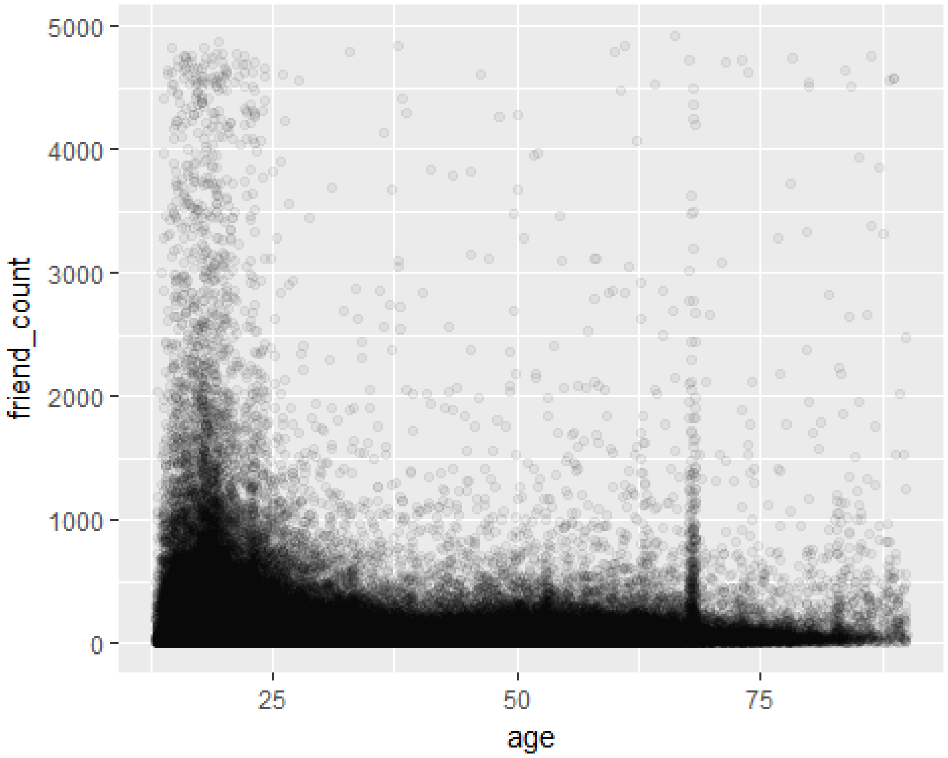
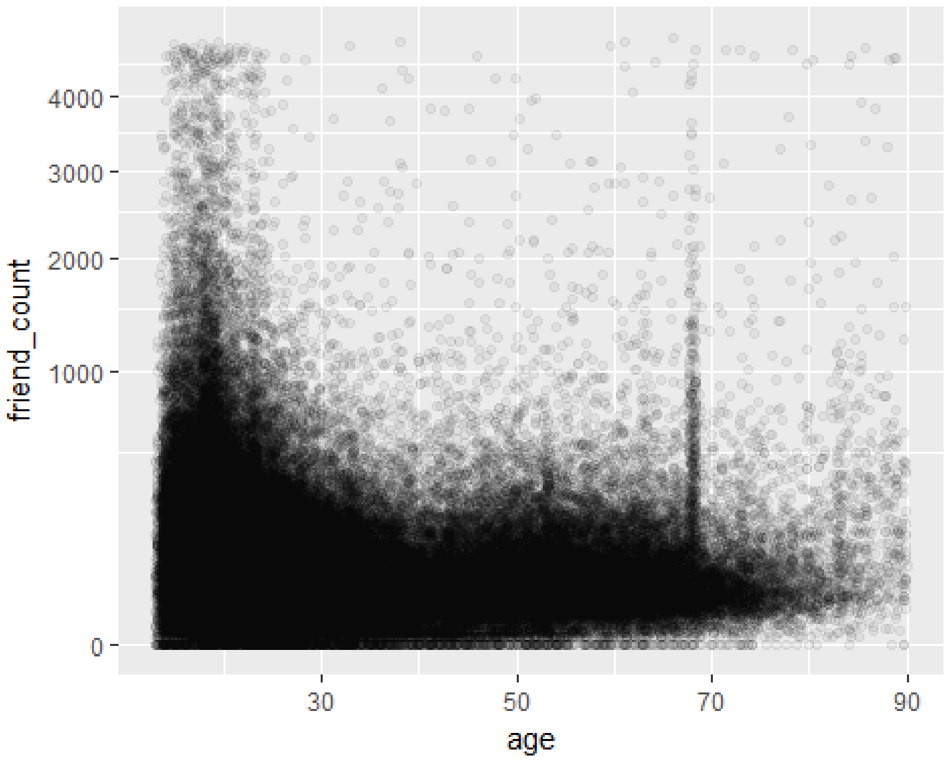
Scatterplots

```
library(ggplot2)
library(dplyr)

##
## Attaching package: 'dplyr'

## The following objects are masked from 'package:stats':
##
##     filter, lag

## The following objects are masked from 'package:base':
##
##     intersect, setdiff, setequal, union

summary(pf$age)

##    Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##   13.00   20.00   28.00   37.28   50.00  113.00
```

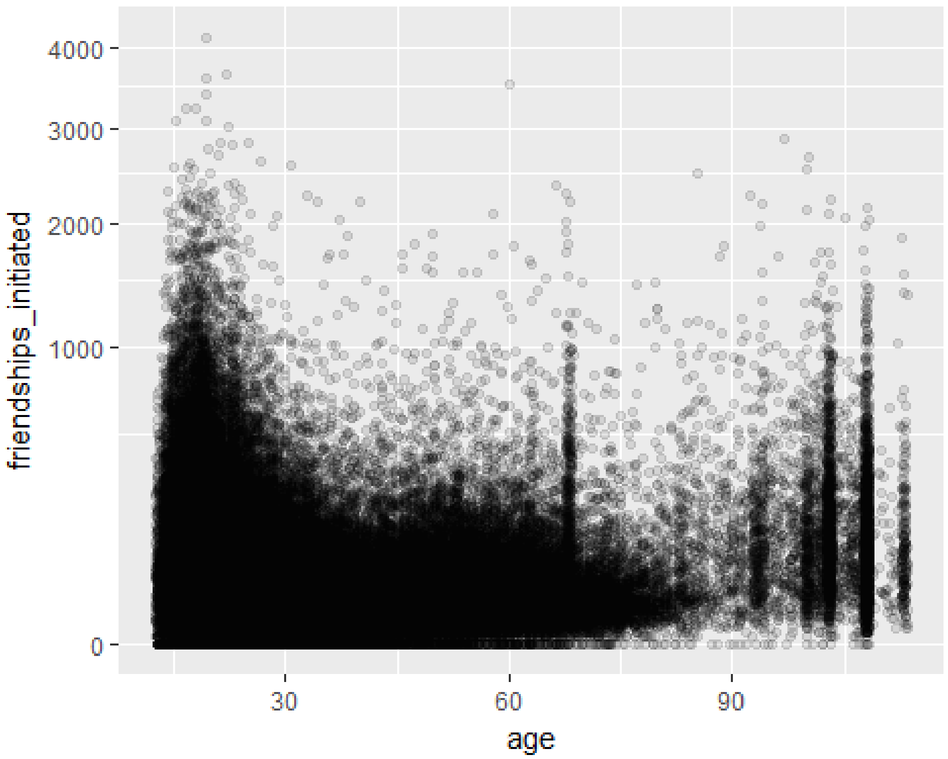
```
#qplot(age, friend_count, data=pf)
#The age variable becomes discrete so we use alpha 1/20 so we can have depth
and understand the numbers
ggplot(aes(x=age, y=friend_count), data=pf) +
  geom_jitter(alpha = 1/20) +
  xlim(13,90)

## Warning: Removed 5188 rows containing missing values (geom_point).
```



```
#Note we need to make sure the minimum height of the jitter to be 0 because w
e're using sqrt
ggplot(aes(x=age, y=friend_count), data=pf) +
  geom_jitter(alpha = 1/20, position=position_jitter(h = 0)) +
  xlim(13,90)+
  coord_trans(y="sqrt")

## Warning: Removed 5175 rows containing missing values (geom_point).
```

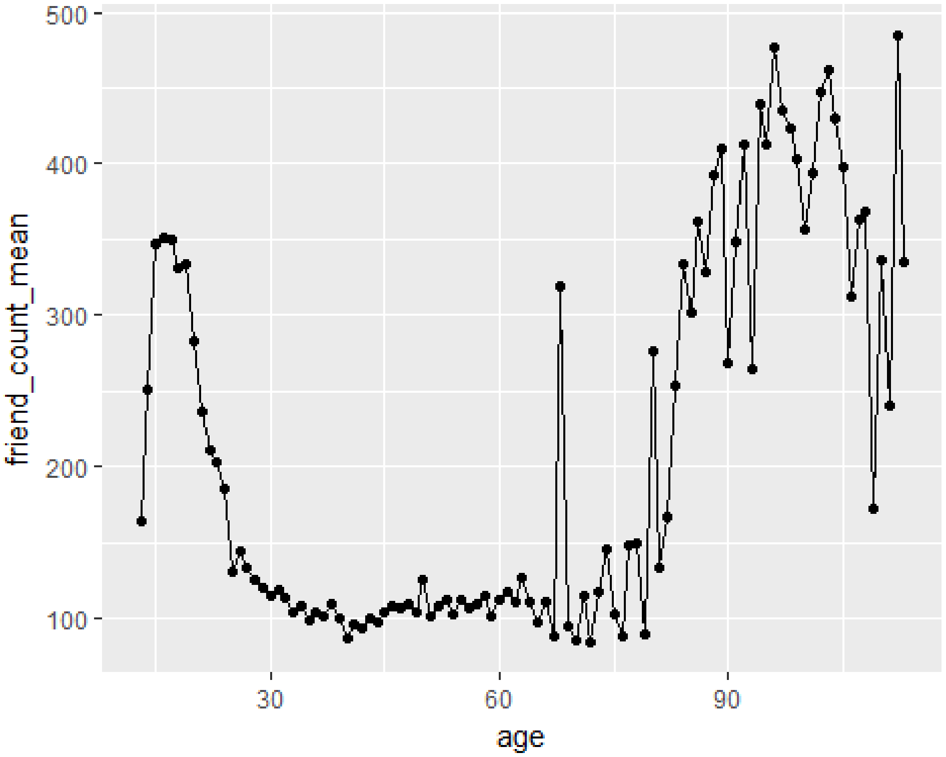


```
ggplot(aes(x=age, y=friendships_initiated), data=pf)+
  geom_jitter(alpha = 1/10, position = position_jitter(h = 0))+
  coord_trans(y='sqrt')
```



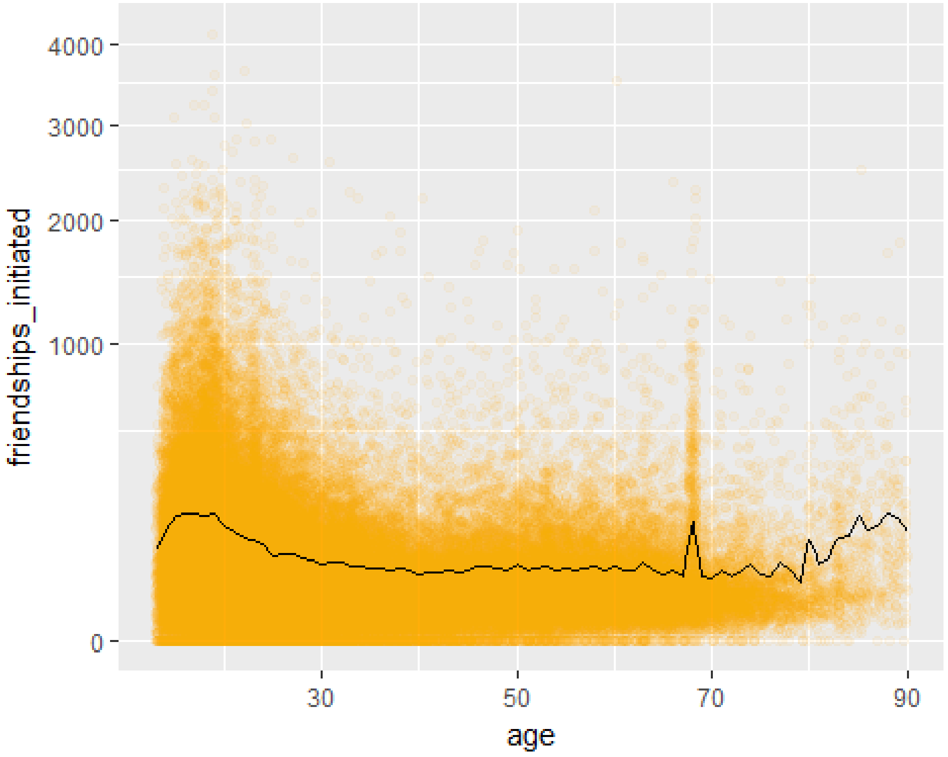
```
#Arranging by age group. The %>% chains the function to the pf dataset
pf.fc_by_age <- pf %>%
  group_by(age) %>%
  summarise(friend_count_mean = mean(friend_count),
            friend_count_median = median(friend_count),
            n = n()) %>%
  arrange(age)
```

```r
#Plotting the median friend count
ggplot(aes(x=age, y=friend_count_mean), data = pf.fc_by_age)+
  geom_point() + geom_line()
```



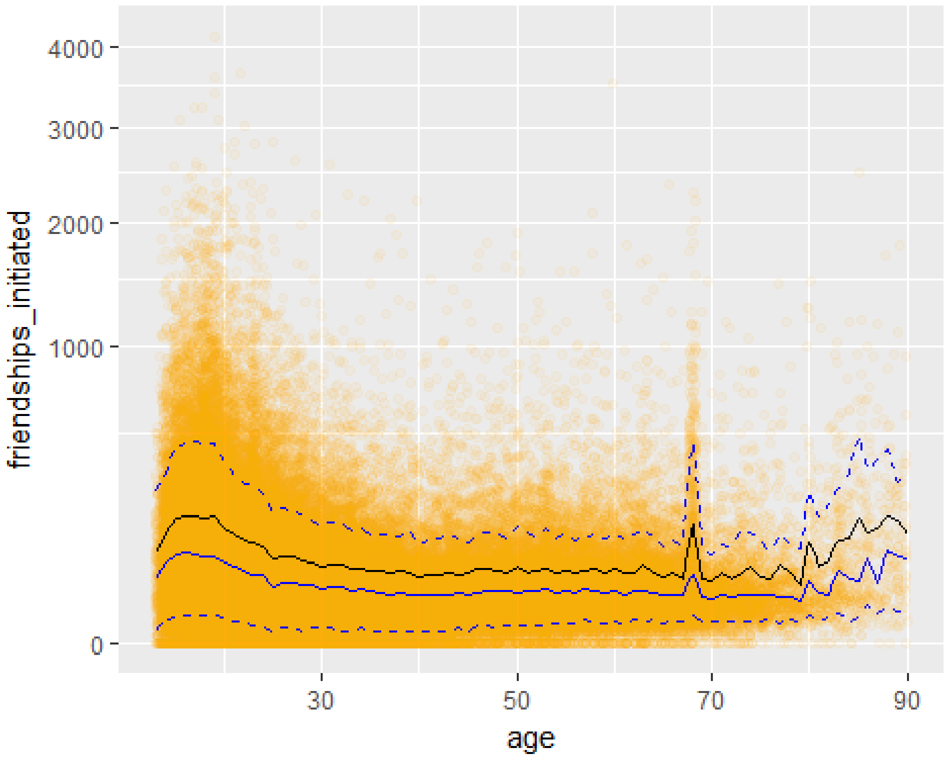
```
ggplot(aes(x=age, y=friendships_initiated), data=pf)+
  xlim(13,90) +
  geom_jitter(alpha = 0.05,
              position = position_jitter(h = 0),
              color = 'orange') +
  coord_trans(y='sqrt')+
  geom_line(stat = 'summary', fun.y=mean)
```

```
## Warning: Removed 4906 rows containing non-finite values (stat_summary).

## Warning: Removed 5181 rows containing missing values (geom_point).
```

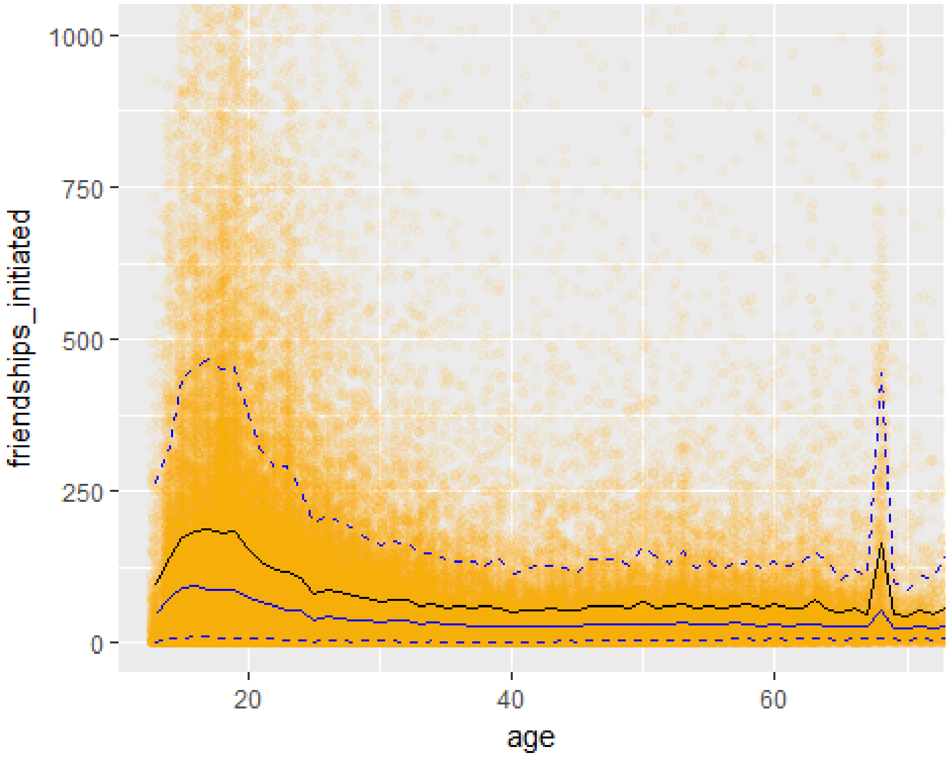


```
#Below we added the have the 90% and 10% quantiles (dotted blue lines)
#The mean is the black line
#The median is the solid blue line
ggplot(aes(x=age, y=friendships_initiated), data=pf)+
  xlim(13,90) +
  geom_jitter(alpha = 0.05,
              position = position_jitter(h = 0),
              color = 'orange') +
  coord_trans(y='sqrt')+
  geom_line(stat = 'summary', fun.y=mean) +
  geom_line(stat = 'summary', fun.y=quantile, fun.args = list(probs = .1), li
netype = 2, color = 'blue')+
  geom_line(stat = 'summary', fun.y=quantile, fun.args = list(probs = .5), co
lor = 'blue')+
  geom_line(stat = 'summary', fun.y=quantile, fun.args = list(probs = .9), li
netype = 2, color = 'blue')
```

```
## Warning: Removed 4906 rows containing non-finite values (stat_summary).

## Warning: Removed 4906 rows containing non-finite values (stat_summary).

## Warning: Removed 4906 rows containing non-finite values (stat_summary).

## Warning: Removed 4906 rows containing non-finite values (stat_summary).

## Warning: Removed 5197 rows containing missing values (geom_point).
```



```
#Using coord_cartesian to zoom in to 600 and below
ggplot(aes(x=age, y=friendships_initiated), data=pf)+
  geom_point(alpha = 0.05,
             position = position_jitter(h = 0),
             color = 'orange') +
  coord_cartesian(xlim = c(13,70), ylim = c(0,1000))+
  geom_line(stat = 'summary', fun.y=mean) +
  geom_line(stat = 'summary', fun.y=quantile, fun.args = list(probs = .1),
            linetype = 2, color = 'blue')+
  geom_line(stat = 'summary', fun.y=quantile, fun.args = list(probs = .5),
            color = 'blue')+
  geom_line(stat = 'summary', fun.y=quantile, fun.args = list(probs = .9),
            linetype = 2, color = 'blue')
```


friendships_initiated
age