

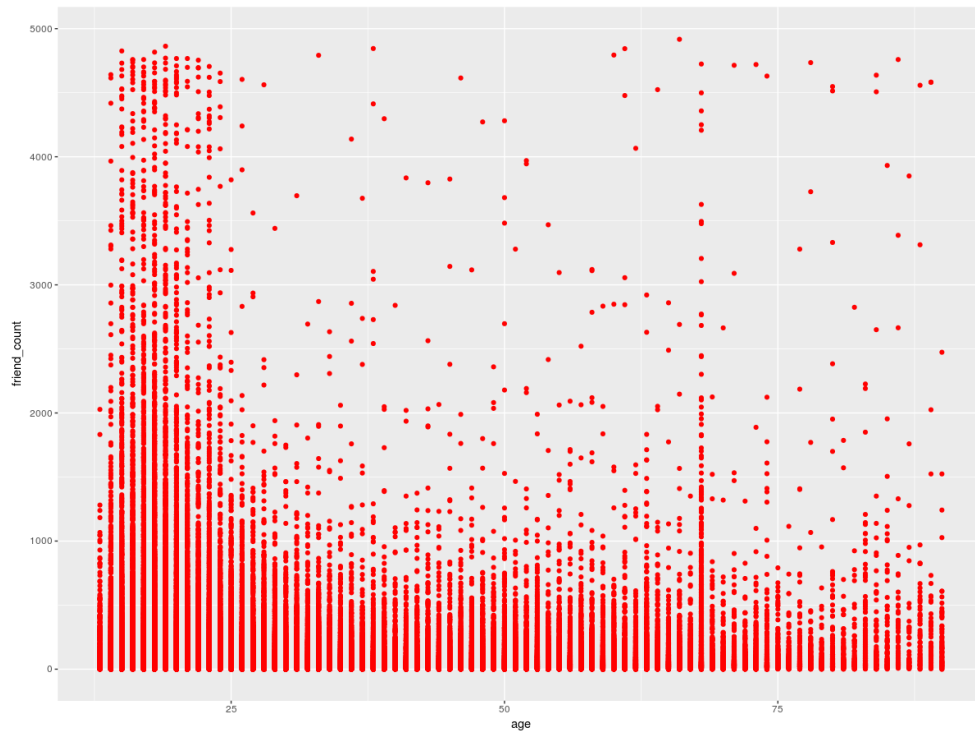
Revising Codes:

Lesson 2 : Exploring 2 variables

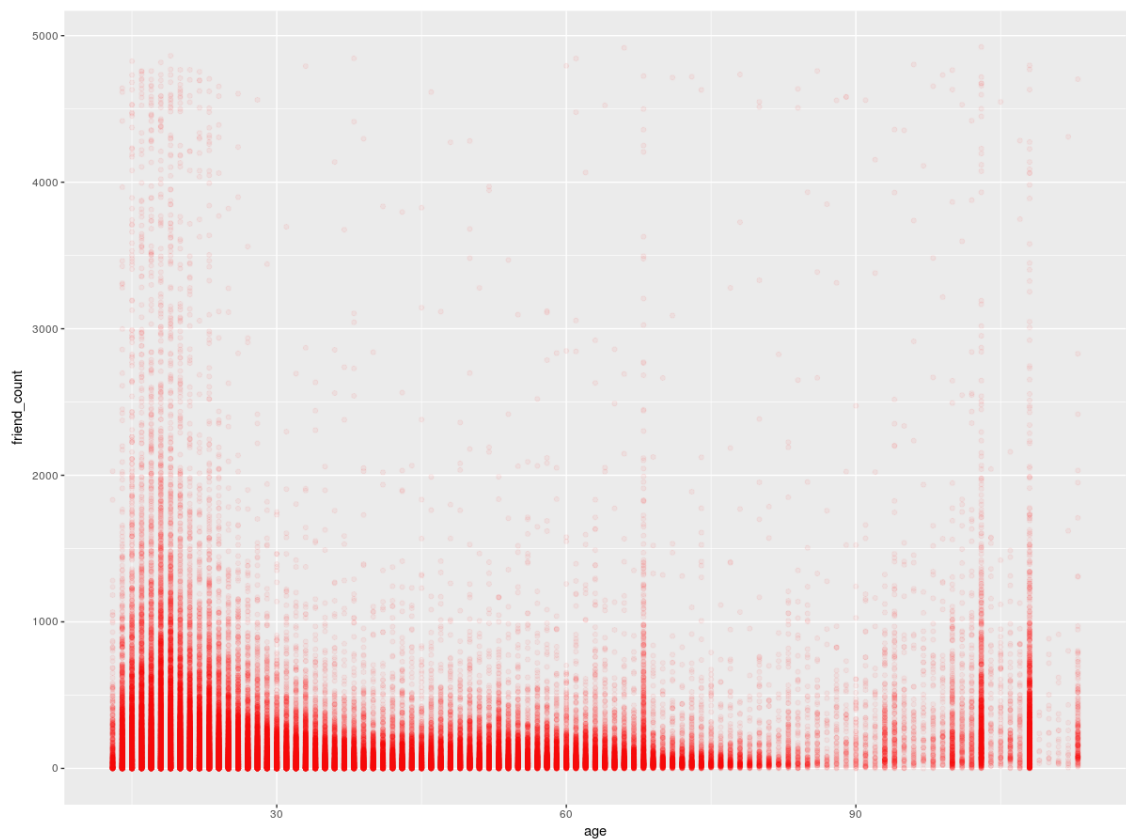
Perform below given tasks using R

Solution plots are given for reference to see if you ever go wrong while practicing.

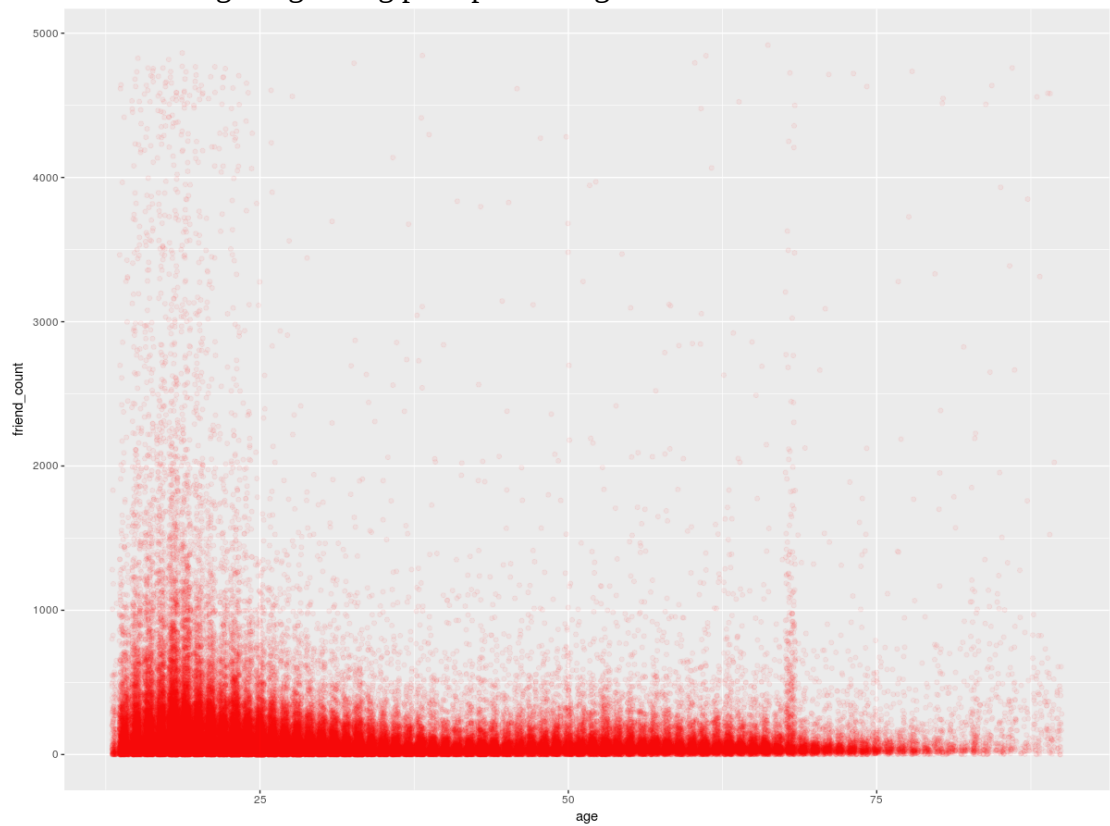
1. Load pseudo facebook data and make scatter plot of friend count vs age. With age range from 13 to 90.



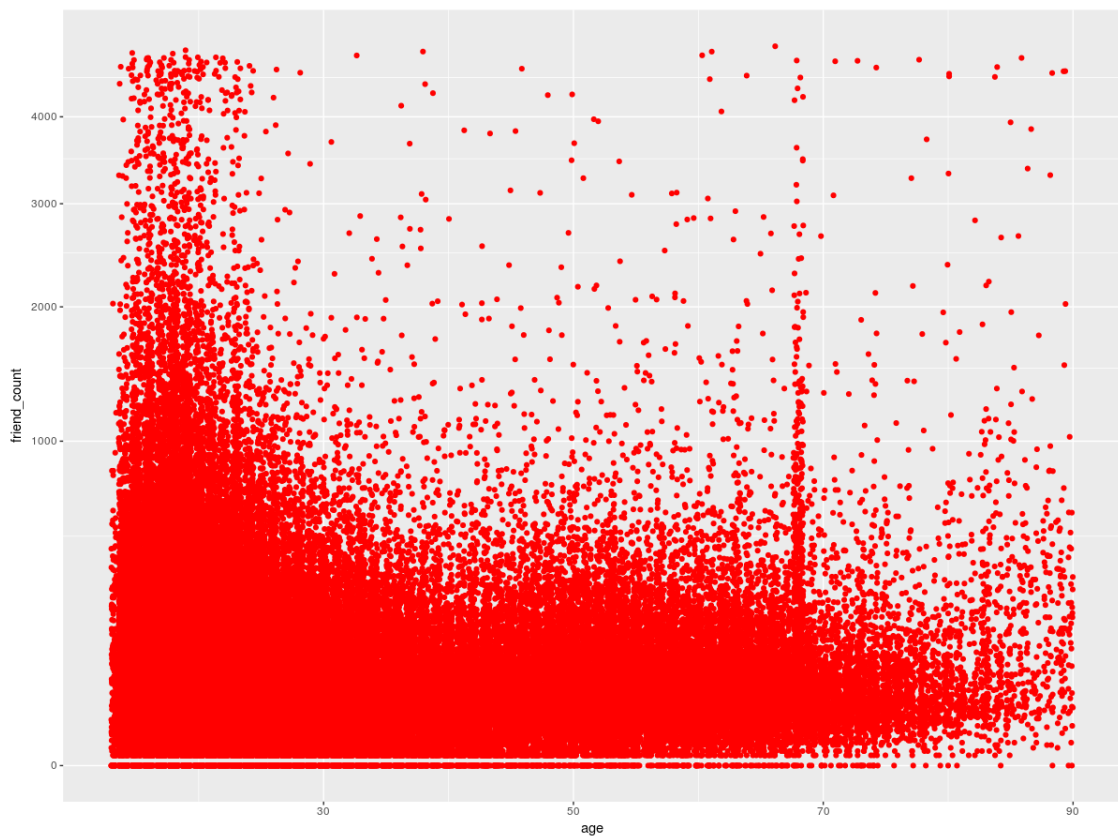
2. Make the same plot again but take care to avoid overplotting.



With some noise to avoid getting wrong perception of age



3. Add conditional transformation layer on y-axis to have better visualization.
Also make condition over jitter to avoid getting wrong results when someone has 0 friend count.



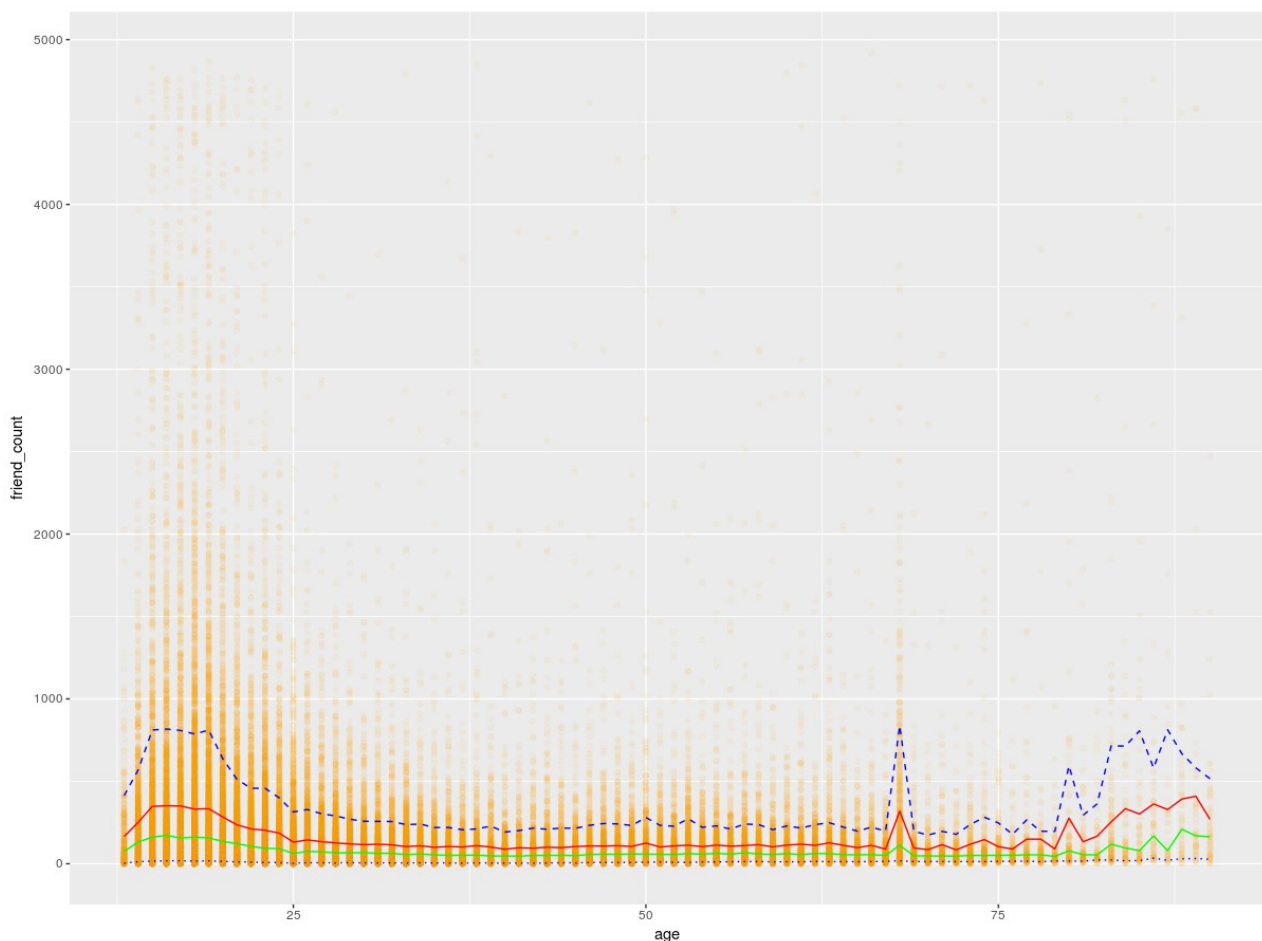
4.

Time to summarise relationship between two variables apart from plotting every single point. For example you may want to find how does average friend count vary over age. We can do this by making a table that will tell for each age mean and median friend count. That's when we use 'dplyr' package in R.

Make a new data frame that will show mean friend count, median friend count and number of people for different ages.

	age	fmeanc	fmedianc	n
1	13	164.75000	74.0	484
2	14	251.39013	132.0	1925
3	15	347.69213	161.0	2618
4	16	351.93714	171.5	3086
5	17	350.30064	156.0	3283
6	18	331.16628	162.0	5196
7	19	333.69210	157.0	4391
8	20	283.49907	135.0	3769
9	21	235.94116	121.0	3671
10	22	211.39479	106.0	3032

5. Make a scatter plot of friend count vs age and line plots of mean friend count, median friend count, 10% quantile and 90% quantile as summary plots over the same scatter plot.



6. Find the correlation coefficient between age and friend count using Pearson method.
-0.02