**Exploratory Data Analysis**



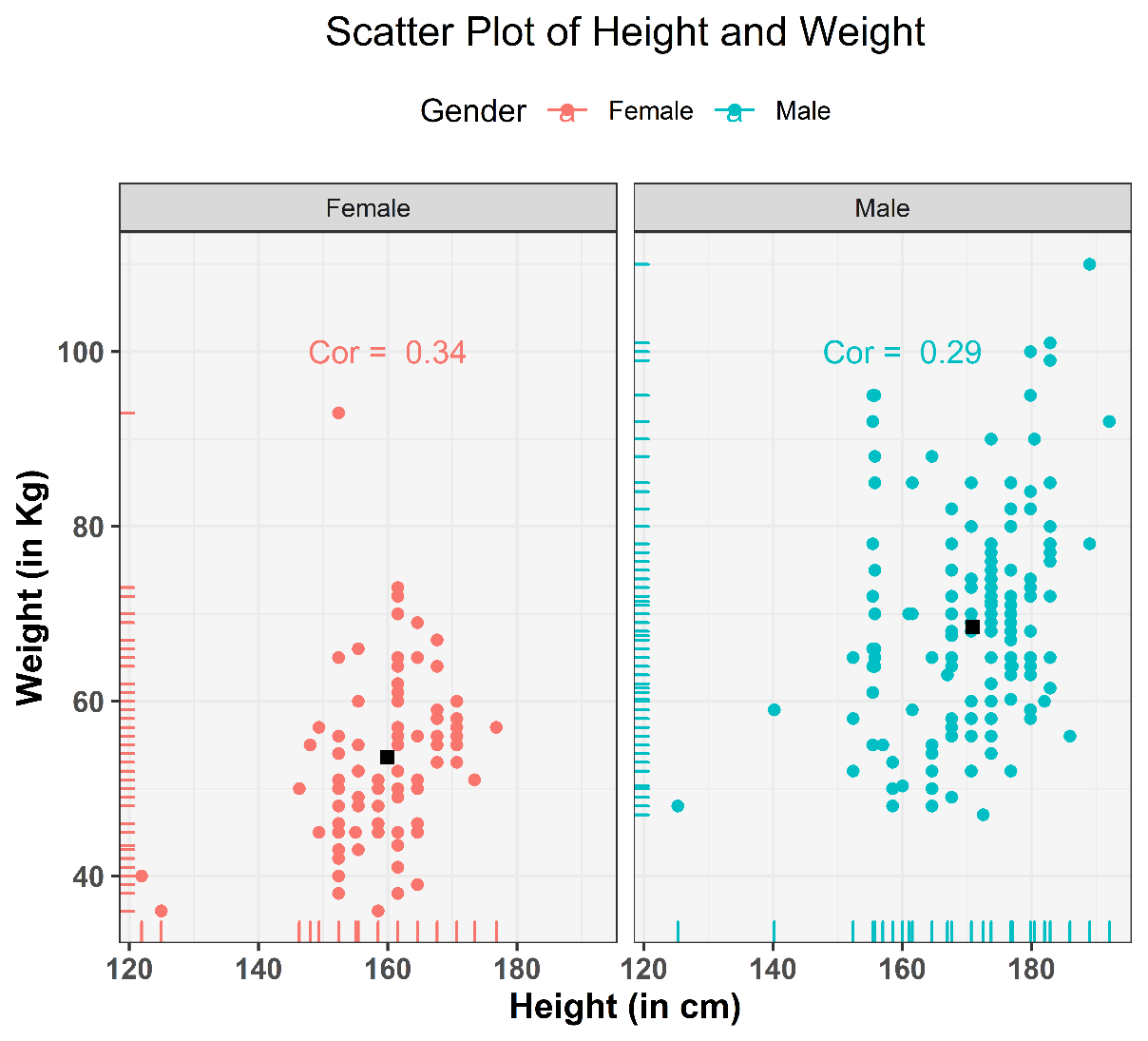
Figure : Age distribution of respondents

The above diagram shows nothing but the ages of our respondents. Now we are categorising the respondents according to their genders.

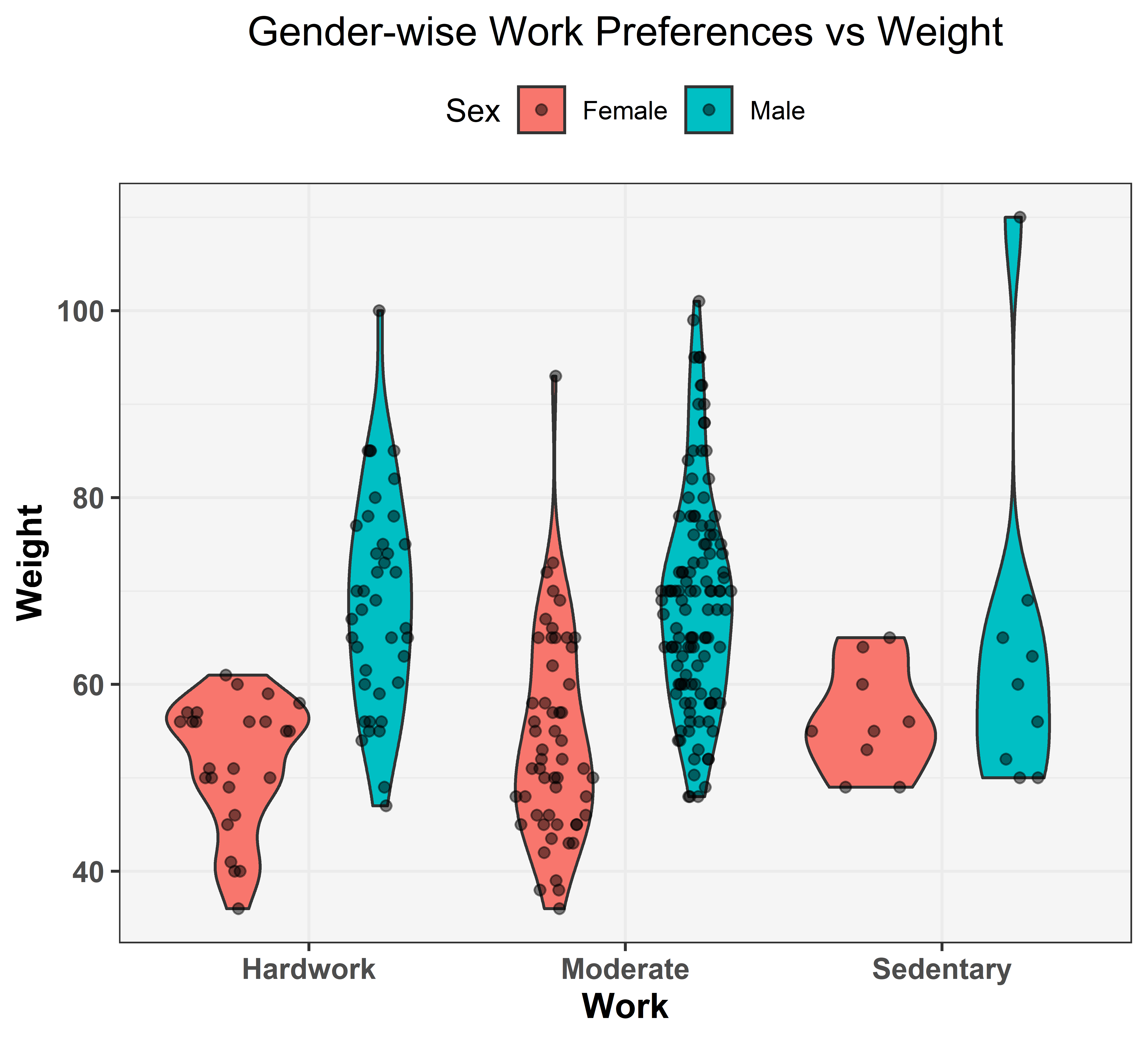
As we can see from the diagram all of our respondents are either male or female. Ages of the female respondents are marked red and the male ones are marked green.

Some other observations are:

* There are total 3 outliers in the dataset, 2 of them are male and the other one is a female.
* Most of the data points lie in between 15 to 25 that means most of the respondents are aged in between 15 to 25.
* 35% of the total responses are from female and 65% are from male that means male respondents are almost double of the female respondents in this dataset.

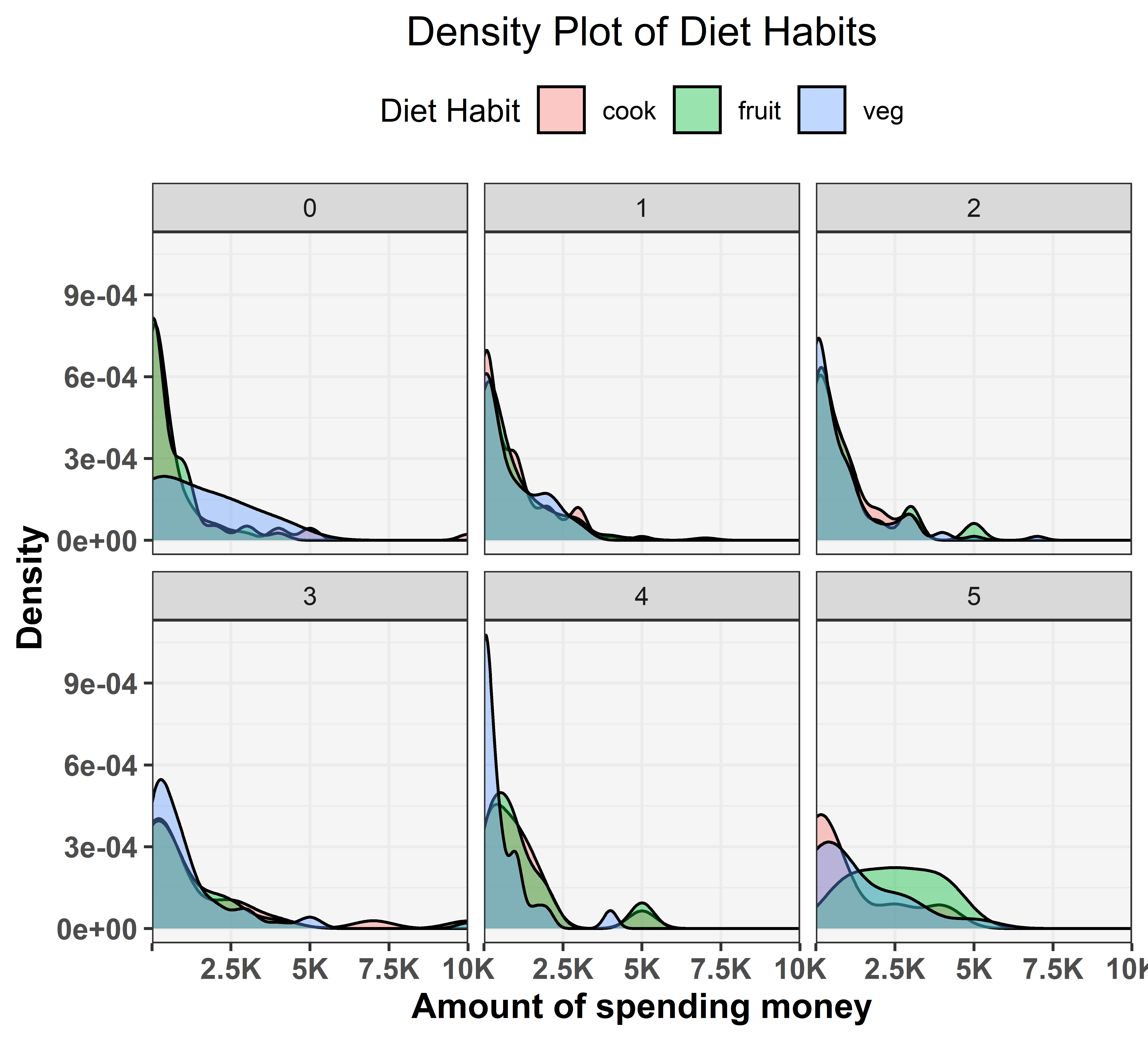


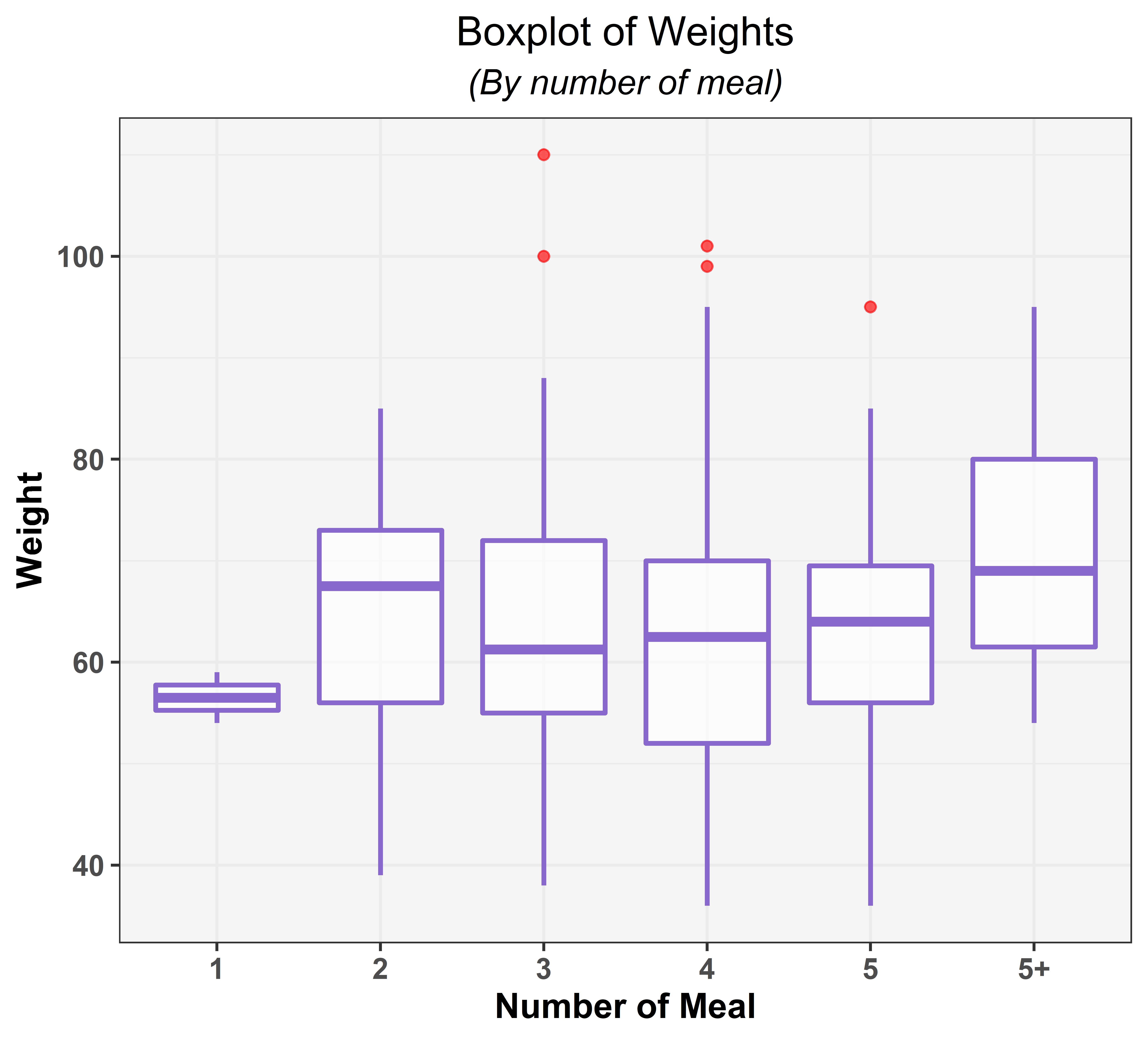
* It is clear that weight of female respondents are mostly in between 40kg to 70kg and height in between 145 cm to 175cm ; for males it is mostly in between 46 kg to 100kg and 155cm to 185cm respectively.
* Surprisingly, weights of males are scattered are more uniformly than females.
* Overall male respondents has more height and weight than females.
* From the correlations values, female’s height and weights are highly correlated than males’.
* From the above graph it is evident that the correlation between height and weight is significantly higher for females as compared to their male counterparts. It is 0.34 for females as compared to 0.29 for males. It is supported by the popular belief that females tend to take care of their body more sincerely as opposed to their male counterparts.



The above diagram is a violin plot of "Gender-wise work preference vs weight of the respondents". Violin plots are quite similar to the Box-plot with the addition of rotated density-plot on each side. Here the red area denotes the female responses and the blue area denotes the male responses. The other observations are:

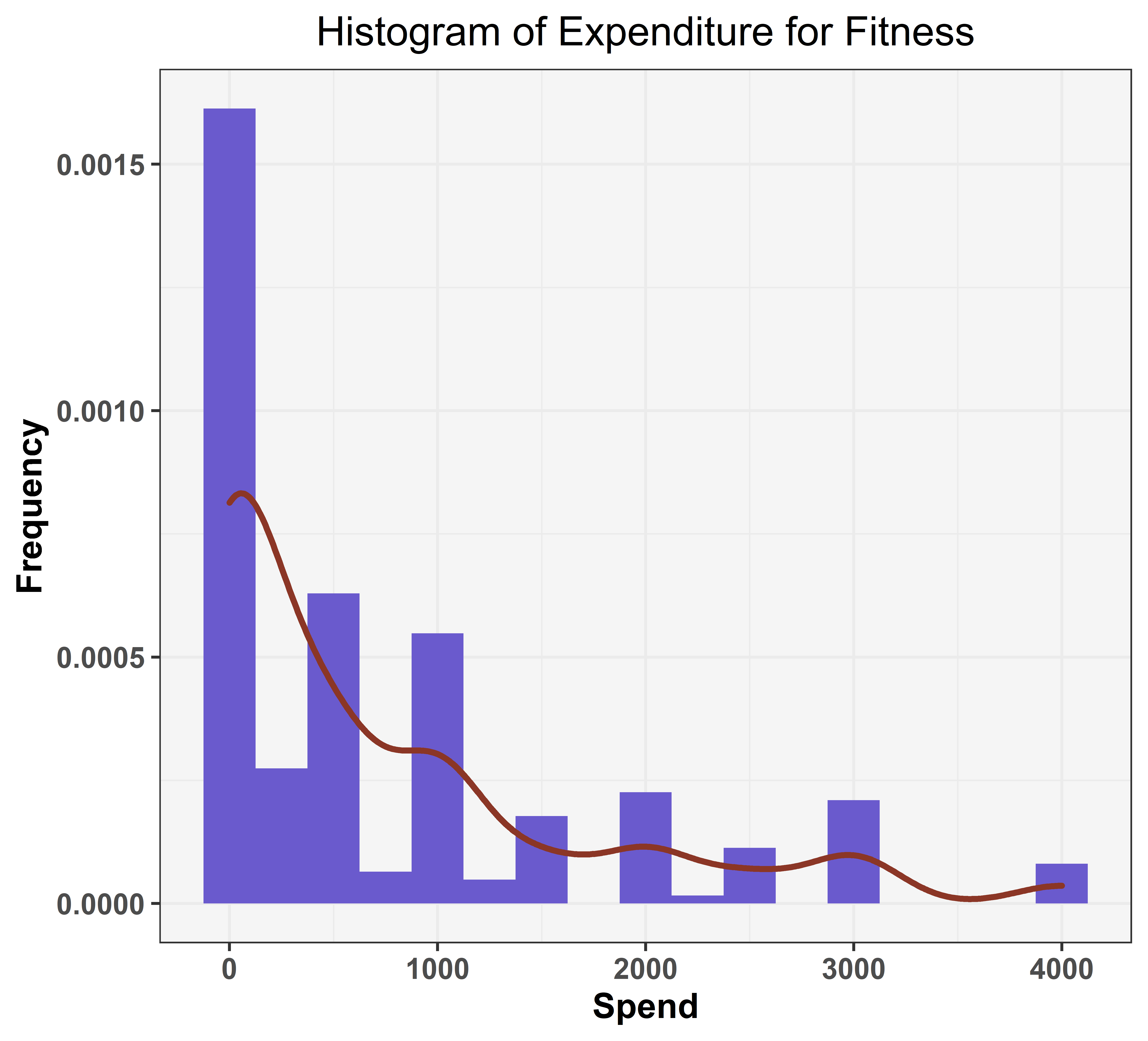
* Most of the hardworking females are weighted between 40 to 60 kg whereas most of the sedentary females are weighted between 50 to 65 kg. But as we can see moderate working females' weights are distributed fairly in between 35 to 80 kg with a bit more density around 45 to 55 kg.
* Similarly if we try to interpret the male respondents' weights according to their work preference, most of the hardworking males' weights are almost fairly distributed between 50 to 100 kg with a bit more density around 60 to 80 kg. Moderate working males are weighted in between 50 to 100 kg in which case their weights are a bit more dense around 60 to 75 kg. Now we have an outlier in our dataset of male responses who is a sedentary person. Other than him, the rest of the sedentary males are weighted in between 55 to 80 kg.
* Now the graph tells us that most of the respondents in our dataset are moderate-workers.



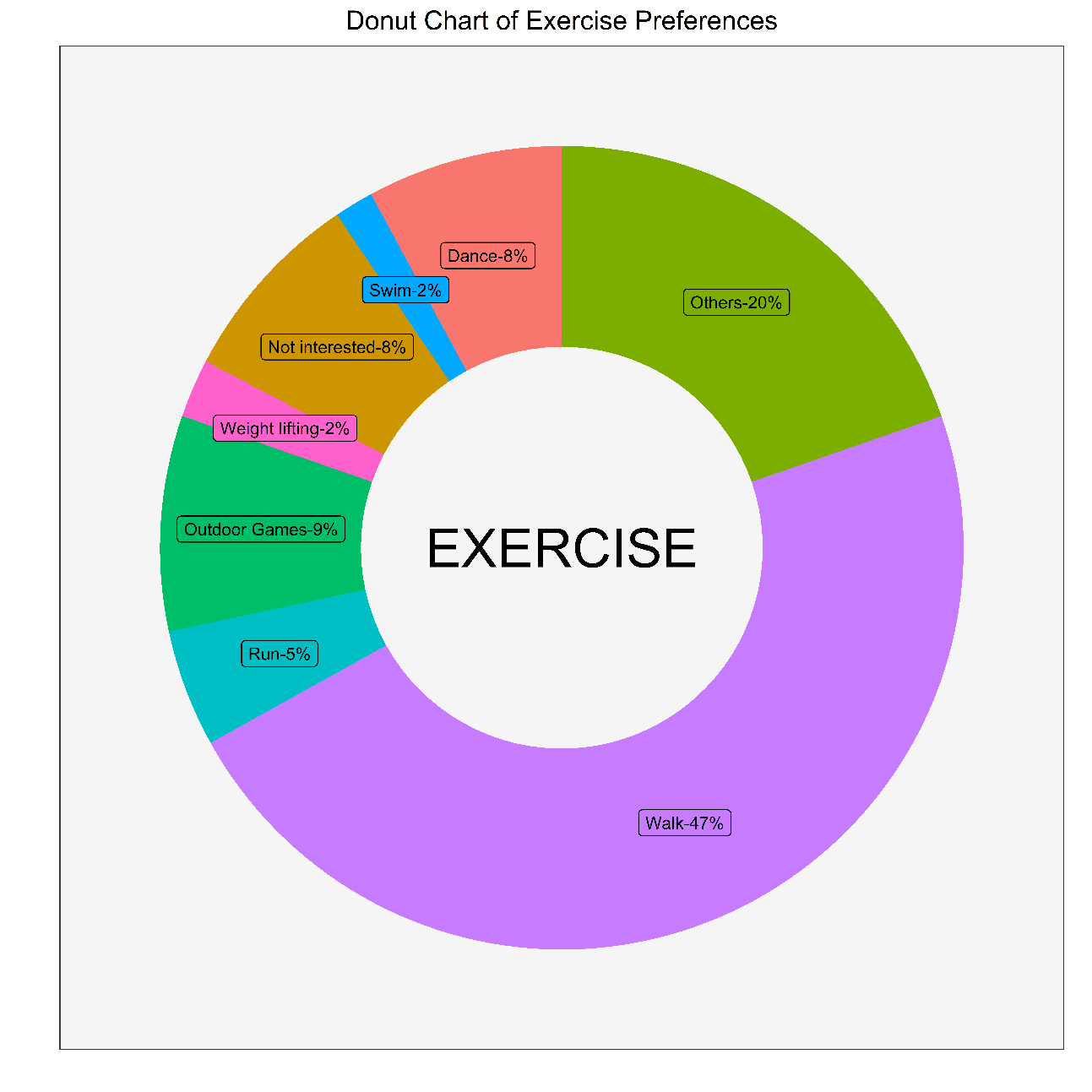


The above diagram is the Box-plot of the numbers of meals taken by the respondents in a day vs their weights. The red dots are some outliers in the dataset. Now some other observations regarding the boxplots are:

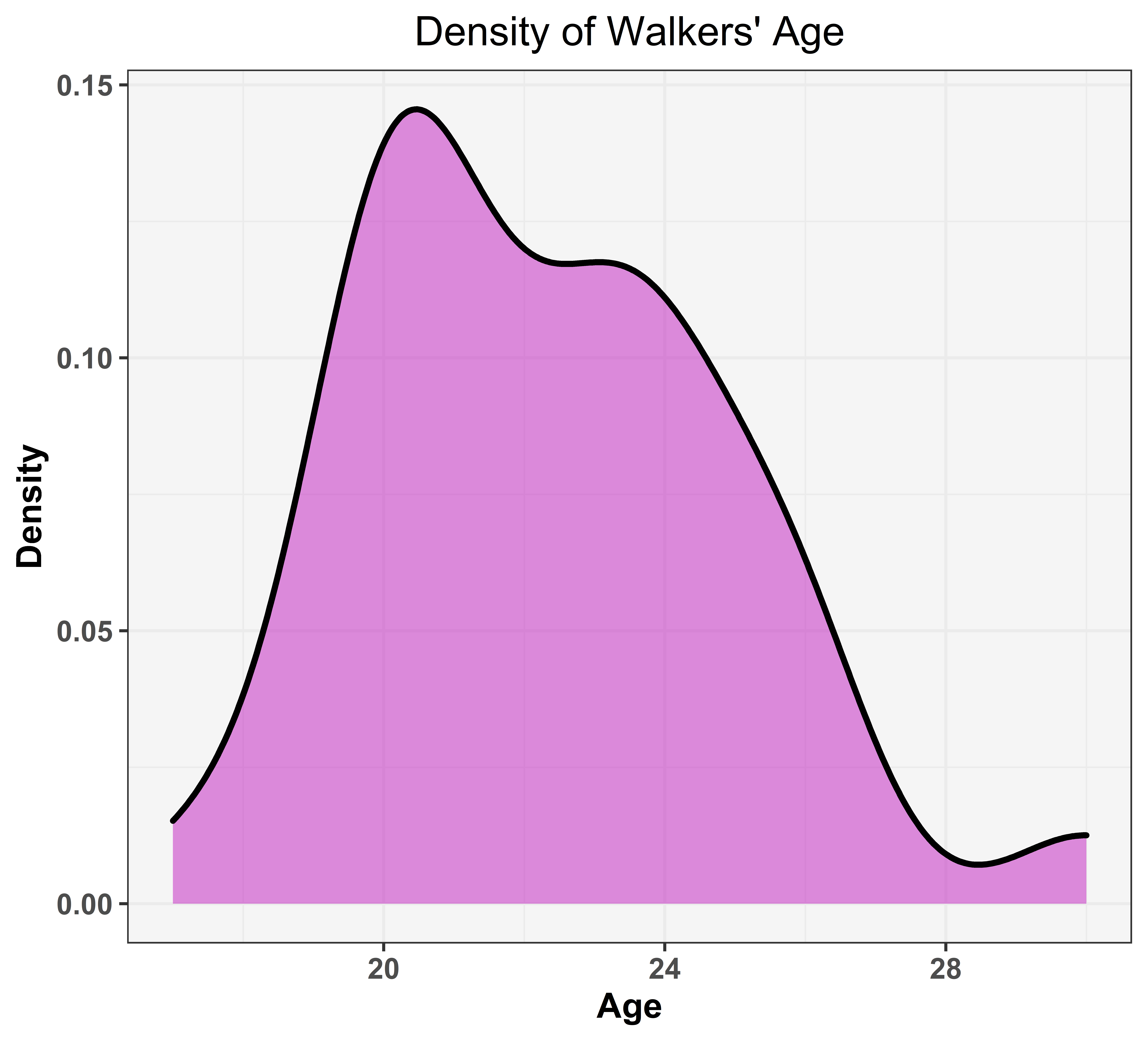
* The respondents who take 1 meal per day are weighted in between 57 to 59 kg with median 58 kg.
* Those who take 2 meals per day are weighted in between 58 to 73 kg with median around 68 kg, which means the weights are denser in between 68 to 73 kg.
* Those who take 3 meals per day are weighted almost similar to the persons who take 2 meals but their weights are denser in between 57 to 61 kg.
* Those who take 4 meals per day are weighted in between 52 to 70 kg with median at 63 kg and also the weights are distributed in that range quite fairly.
* Those respondents' weights who take 5 meals per day are in between 56 to 70 kg and the median weight is 64 kg. Those who take more than 5 meals per day are weighted in between 62 to 80 kg and the median weight is 69 kg.
* As we can see most of the respondents who consumes more than 5 meals per day are heavier than others which is quite natural. The other 4 categories (2, 3, 4, 5 no. of meals takers) are weighted quite similar. Two of the five outliers who take 3 meals per day are weighted 100 kg and 110 kg, two of them who take 4 meals are weighted around 100 kg and the other one takes 5 meals per day is weighted 95 kg.



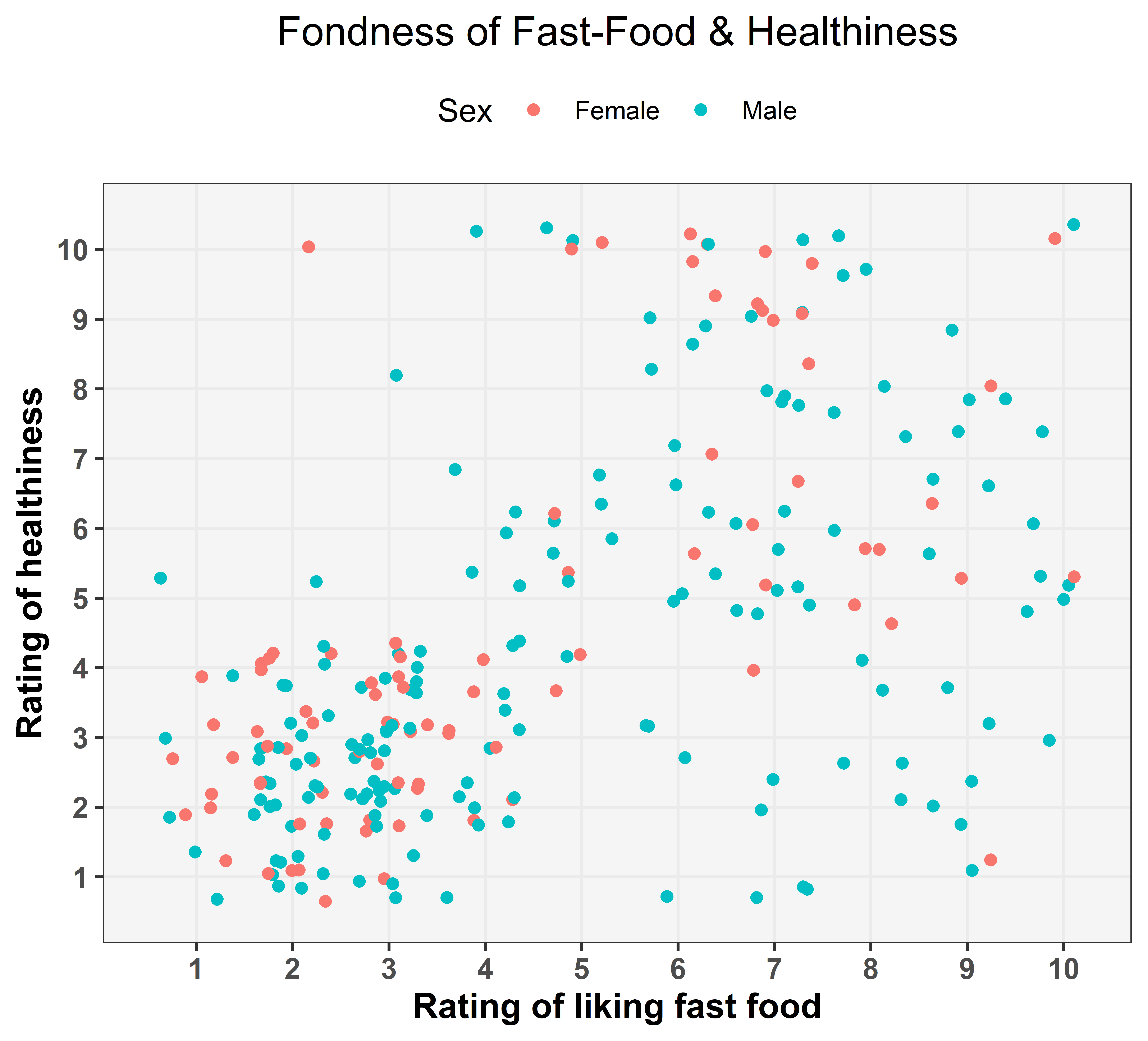
The above diagram shows that most of the respondents either do not pay any money or pay less than 1000Rs. on fitness. The density curve decreases as the value of money spent on fitness increases that means no. of person spending more and more on their fitness decreases gradually.



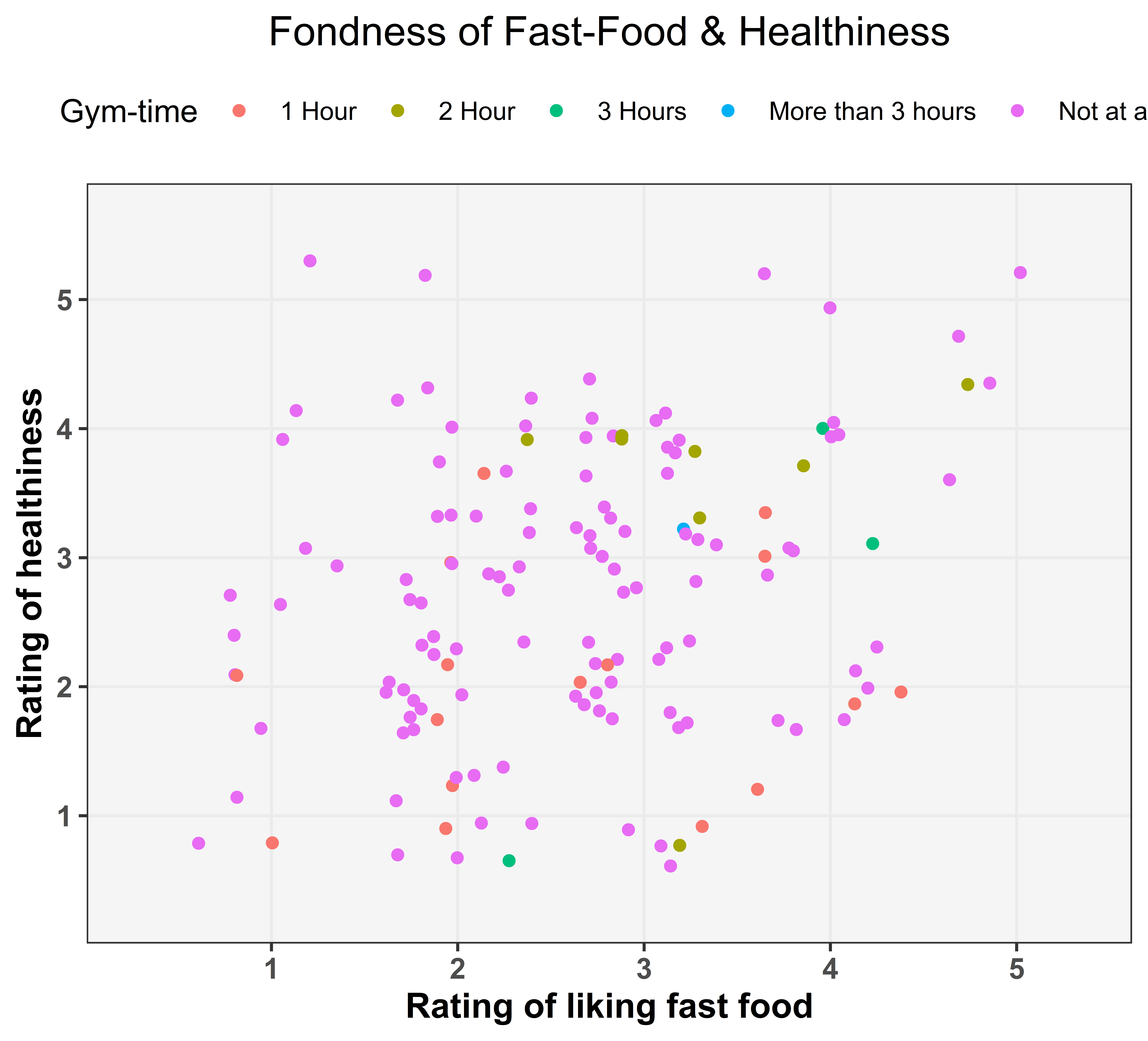
From the above graph, we can infer that 47% of the people have preferred walking as the primary mode of exercise while the second most preferred form of exercise is not known as 20% of the people have said others.

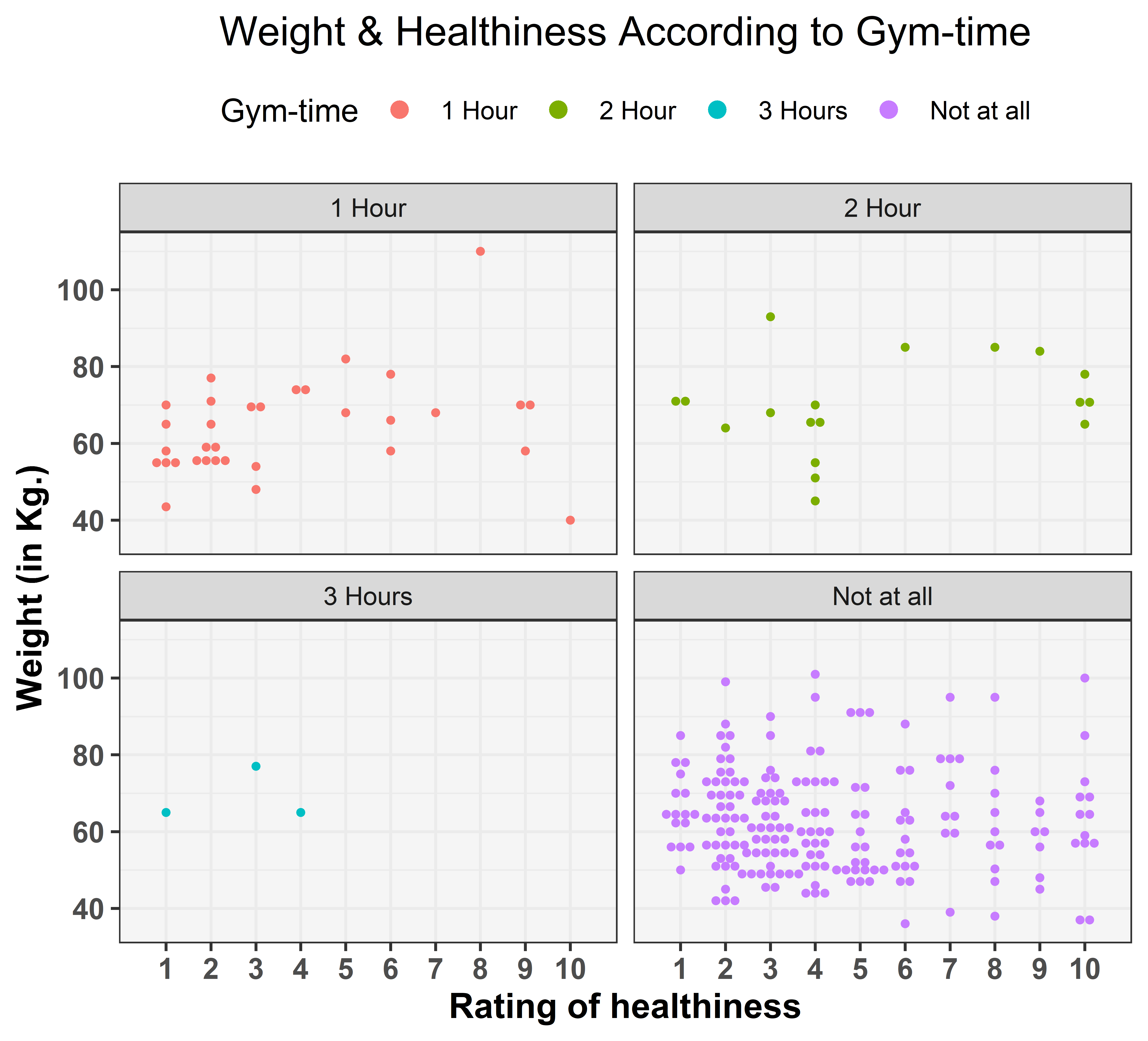


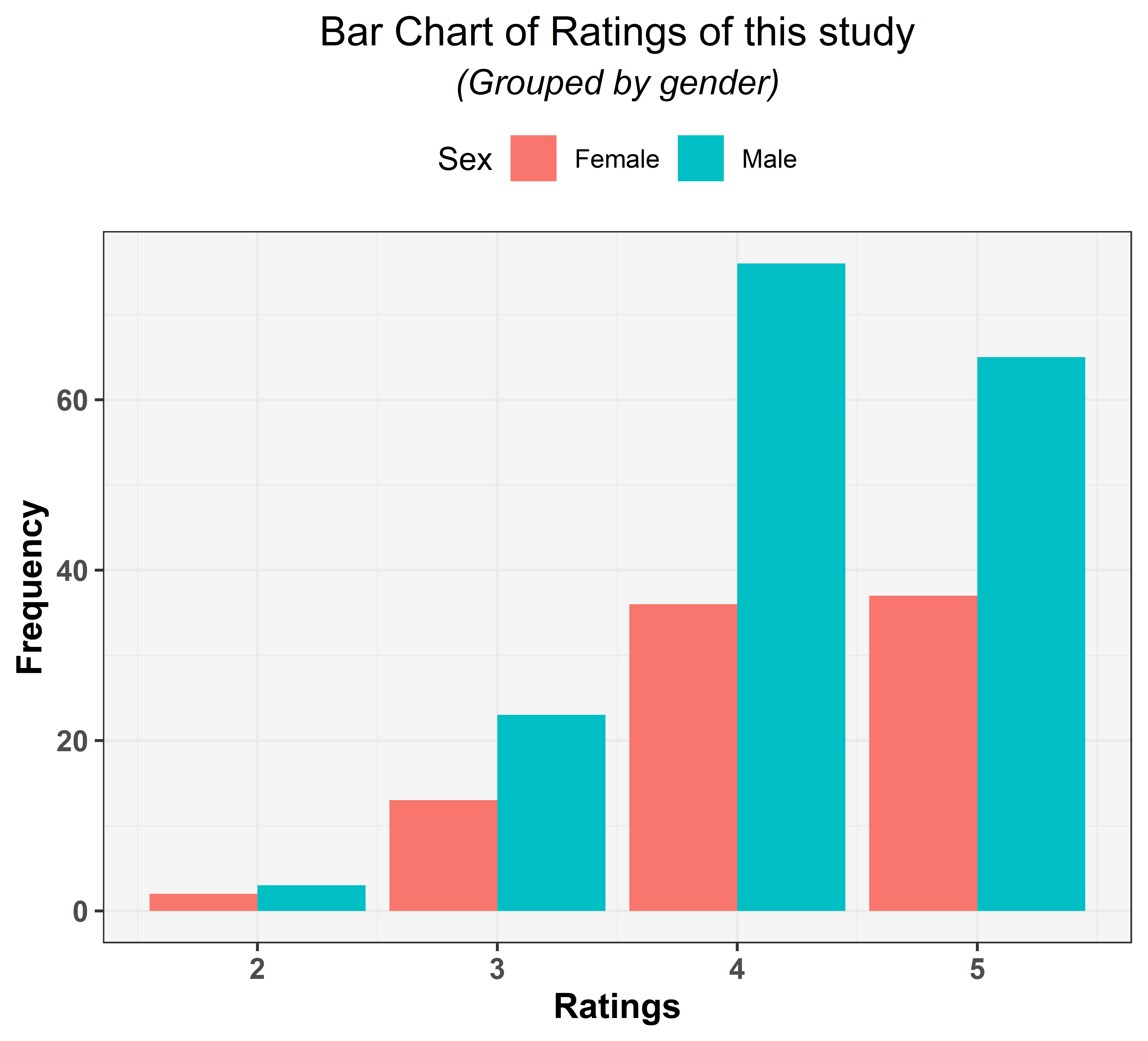
The age distribution of the respondents who have preferred walking as their primary mode of exercise have been given below. As opposed to the popular belief that young adults are more inclined towards intense exercises such as running or weight lifting, the respondents of walking have the age mostly in the range of 18-22!



The above graph has suggested us with the finding that there is group cluster of males and females who have not rated their physical health very high despite not having high likability for fast food. The reason for this can be well understood from the graph below where we have linked this cluster's health parameter with the help of time spent in gym by these people.

When we look closely at the dataset of the above cluster with respect to gym time we have found the rationale behind the unusual low rating of health despite not giving high rating for fast food preference and that is, most of the people making up the that cluster have reported that time spent in gym -"not at all". This might be a probable reason for this cluster behaviour.





As per the above graph, we can see that we have received more responses from the males for the positive ratings as compared to the female counterparts. Also out of the total responses x % have given us a rating above 4.