

# **A CASE STUDY ON VACCINATION** **IN INDIA**

**(EXPLORATORY DATA ANALYSIS USING R AND  
PYTHON)**

**NAME - UDVAS DAS & NILABJA GHOSH**

**RAMKRISHNA MISSION VIVEKANANDA EDUCATIONAL AND  
RESEARCH INSTITUTE**



## **DECLARATION**

“We affirm that we have identified all our sources and that no part of our project work uses unacknowledged materials.”

Signed,

Udvas Das

Nilabja Ghosh

# Contents

INTRODUCTION .....	3
ABSTRACT .....	4
DATA COLLECTION .....	5
LOOKING AT THE DATA .....	6
DEMOGRAPHICAL VISUALIZATION .....	8
DATA VISUALIZATION.....	12
CONCLUSION .....	34
LIMITATIONS OF THIS PROJECT WORK.....	35
ACKNOWLEDGEMENT .....	36
BIBLIOGRAPHY .....	37
APPENDIX.....	38

# INTRODUCTION

On 11<sup>th</sup> March 2020 the World Health Organization (WHO) characterized the outbreak of the SARS virus named Covid-19 as a global pandemic. In the past two years, the entire human race has experienced the most horrific time in the past century. A lot of things have changed since then including our lifestyles, necessities, day-to-day living etc. Our physical health, especially our mental health got affected heavily.

Many interesting projects in Biology, Statistics, Environmental sciences have come to light since the outbreak of the deadly disease. There is a lot of data which has been collected for these projects.

India reported its first covid case on 30<sup>th</sup> January, 2020 and after August the real havoc started. India became the state with second highest case numbers. After a year of immense struggle and hardship India started to vaccinate its citizens.

A lot of people in India were sceptical about vaccine because of the fake news circulated on social media and digital media in general. Vaccination in India progressed very slowly and took huge time to even vaccinate the elderly population. There are many reasons behind it apart from peoples' perspectives. Nearly after a year or so it became possible to start vaccination for young people who are below 18 years age. Most of the young population is still not vaccinated, even with the first dose!

We have taken up this study titled "A case-study on vaccination in India" because we wanted to visualize people's reliability on the vaccines provided in this country by different companies, the different parameters in the vaccination process with respect to the resident area types of residents etc. We were also interested to see if there are any patterns for different age groups, gender and resident area towards the prejudices.

We will dive deep into visualization as we proceed in this project and draw meaningful inferences on the vaccination in India. The study will involve different parameters which are interlinked and the results will be multi-dimensional.

# ABSTRACT

## (PURPOSE OF THE STUDY)

We have taken up this project in the first place because we noticed indifference in people's mind about vaccine around us. Everyday there is a new rumour about vaccines and even about covid that comes up and people start being panicky about it. We have set our priorities straight for this project. The main purposes in which we are focusing are,

- How many people think vaccination is safe and does it depend on different age intervals, gender and resident places of the subjects?
- What is the vaccination status of the subjects who participated in the study and do age, gender and resident area have a role to play in this?
- What is difference in vaccination status of the people who think vaccination is safe versus who feel the opposite?
- On which type of vaccine providing body people are mostly relying – Government or private organisations? Do age and resident of living have any effect on this mindset?
- Which type of resident area shows maximum vaccine availability, that is subjects from which resident area had to travel least to get a vaccine jab?
- How many people used CoWin app or site to book an appointment for their vaccination? What is the behaviour across the different age intervals and resident types regarding use of the app?
- How easy it was for people to book an appointment? Was booking through CoWin app hassle free?
- On which vaccine producing company or brand people are mostly relying on?
- After which dose of vaccine people are suffering from post-jab symptoms? Do the different vaccine brands have different behaviour regarding these symptoms? Are there any behavioural changes with respect to age intervals and gender regarding post-jab symptoms?
- For those people who did not travel to get the vaccine, that is people who took the jab in their resident area, how the different parameters were taken care of when they went to the vaccination centre like - 1. Hygiene, 2. Social distancing, 3. Waiting time in queue
- Also, we aim to find if the columns are correlated for meaningful inferences.

# DATA COLLECTION

## (PROCESS AND SHORTCOMINGS)

The primary step that is involved in the data analysis process is collection of data. In order to that, we need to have a set of questions, to which the selected cohort will respond. We made a questionnaire using google form and circulated online via messaging mediums to reach the subjects. Since, the topic we chose is vaccination, our target population became anyone and everyone from every age groups. The questionnaire<sup>1</sup> you can find in the appendix of this project report. The questionnaire we made involved some demographic questions like age, gender and resident area which will help us to know our subjects more or we can say we can know the composition of our data well. Then, the questionnaire also involves some questions that are more aligned with the project work including questions about their vaccination status, usage of CoWin<sup>2</sup> portal, reliability on vaccines provided by the Govt and other bodies, their symptoms after the jab, choices of vaccine brands, different safety and health measures maintained in the vaccination centre etc.

Though our study cohort every person in India, the data collection has its own shortcomings too. Since, we don't have enough money and manpower, we cannot collect data by going to people personally and record their responses, we had to rely on google forms and sharing it online. That way we could not reach too many people. Again, vaccination of below 18 years is not yet started in full-fledged manner and also it is hard to reach to non-adult young people via social media, so their data is next to redundant. On the other hand, most of the elderly population is not technologically expert, so it has also been hard to collect their data also. There is a geographical constraint also in the study. Though our project has 'India' in it, we could not reach every corner of the country.

Keeping all the difficulties aside we could collect 131 responses in total which is sufficient to give conclusive results. Now we will look at our data to know the variable<sup>3</sup> names and also, we will consider the demographical angle to know the composition of the subjects who took part in our study.

---

<sup>1</sup> Screenshots of the google form shared

<sup>2</sup> An online portal in India launched by the central Government to book appointment for vaccination

<sup>3</sup> Column names

# LOOKING AT THE DATA

(VARIABLE NAMES AND DEMOGRAPHIC COMPOSITION)

After extracting the data from the google form, the column names were the entire questions by default. So, we changed the names of the columns or variables to make them understandable. The variables or columns involved in the study are given by,

- **AGE**: Gives us the age interval of the subject. Five intervals were set namely, below 18, 18-25, 26-36, 37-45, 45 above.
- **GENDER**: Gives us the gender of the respondent.
- **LIVING AREA**: Tells us the resident type of the respondent, whether it is rural or small city or a metropolitan city.
- **VAC SAFE**: This column records the data if the subjects feel vaccination is safe for their bodies.
- **VAC STATUS**: Tells us the vaccination status of the subject whether he/she is fully or partially or not vaccinated at all.
- **TRAVEL**: This column tells us whether the respondent had to travel somewhere else to get the jab.
- **FIRST CHOICE - SIXTH CHOICE**: This is a (131 x 6) matrix giving us first to sixth choice of vaccine brands for different respondents.
- **VAC BRAND**: The brand of vaccine actually received by the respondent is recorded in this column.
- **VAC RELIABLE**: We stored the information about which type of vaccine provider is preferred by the respondent - Govt. provided free vaccine or paid vaccines provided by the private bodies.
- **COWIN**: This column records if the respondent used CoWin app to book appointment for his/her vaccination.
- **VAC APP EASY**: This gives data on a Likert scale<sup>4</sup> of 1 to 5 telling us how easy it was booking the appointment for vaccination. (1 denotes very easy, 5 denotes very hard)
- **WAITING TIME**: This column represents the experience of the respondents in the vaccination centre regarding the waiting time in queue in hours.

---

<sup>4</sup> This is a linear measuring scale where each value is considered to be equally spaced one after another.

- **SOCIAL\_DIS**: This column represents the experience of the respondents in the vaccination centre regarding the social distancing maintained. This is also represented through a Likert scale from 1 to 5, 1 denotes no satisfaction at all and 5 denotes high satisfaction.
- **HYGIENE**: This column represents the experience of the respondents in the vaccination centre regarding the social distancing maintained through a Likert scale from 1 to 5, 1 denotes no satisfaction at all and 5 denotes high satisfaction.
- **VAC\_SYMPTOM**: We recorded the data on the post vaccination symptoms in this column. It tells us whether the respondent suffered from any symptoms and if yes, then after which dose, he suffered the most.
- **GOVT EFFORT**: We also wanted to know people's opinion about efforts given by the Govt. in supplying enough vaccines and making a safe environment for conducting vaccination process. This was also recorded using a Likert scale from 1 to 5, 1 being that the respondent wasn't convinced enough while 5 being that the person is highly convinced with the effort made by the government.

Now, we will visualize the data demographically<sup>5</sup>.

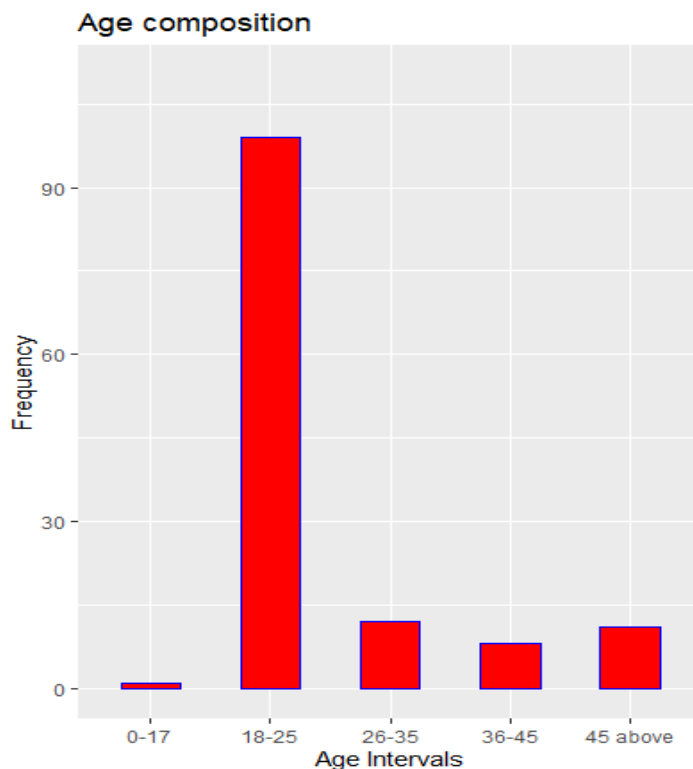
---

<sup>5</sup> Demography means the characteristics associated to an individual which tell us about their age, gender etc

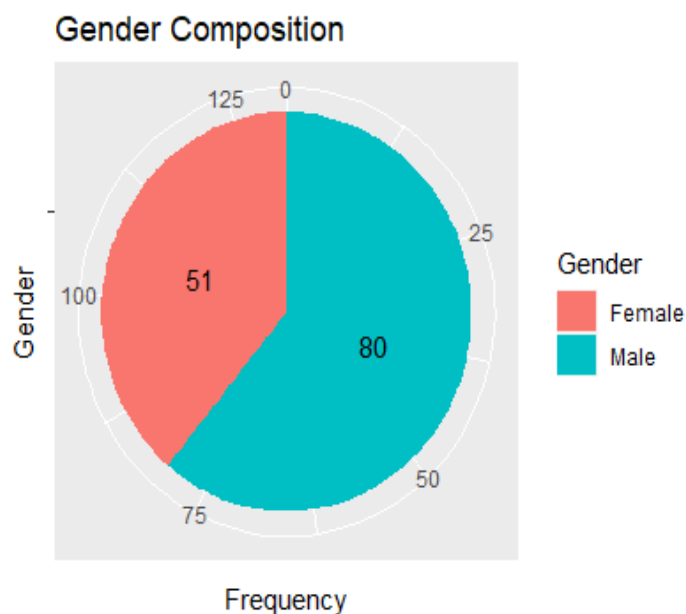
# DEMOGRAPHICAL VISUALIZATION

(DATA COMPOSITION WITH RESPECT TO AGE, GENDER AND RESIDENT AREA TYPE)

When we collect data of a target cohort, we need first to see the composition of the dataset. So, we will look into these things one by one,

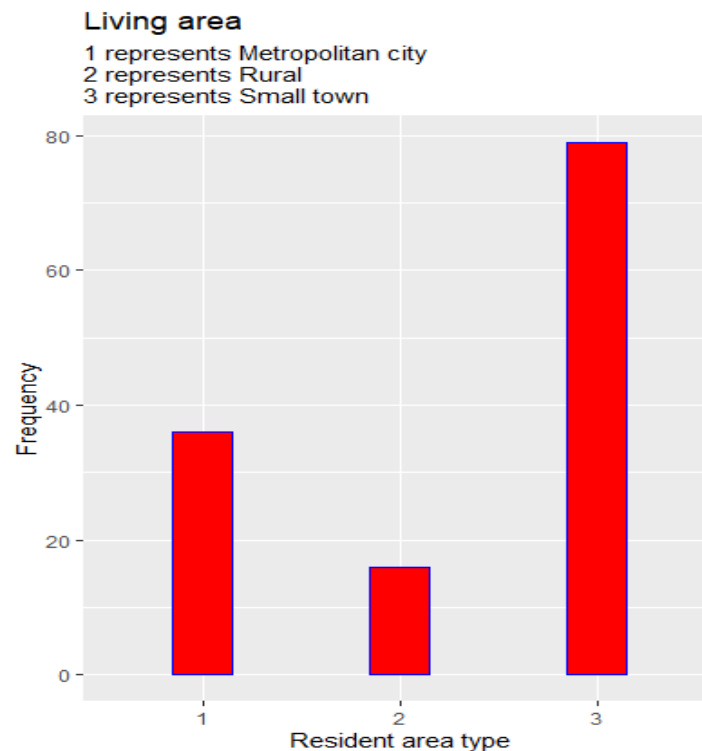


- Most of the respondents in our dataset belong to the age group 18-25, since they were easy to reach.
- We could find lowest number of respondents from the age group below 18.

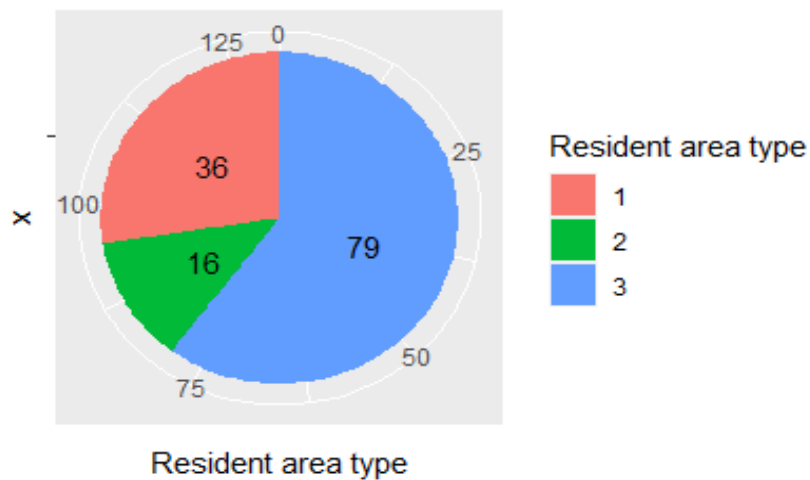


- Number of male respondents in our data is 80.
- Number of female subjects is 51.
- No data for others gender has been received.



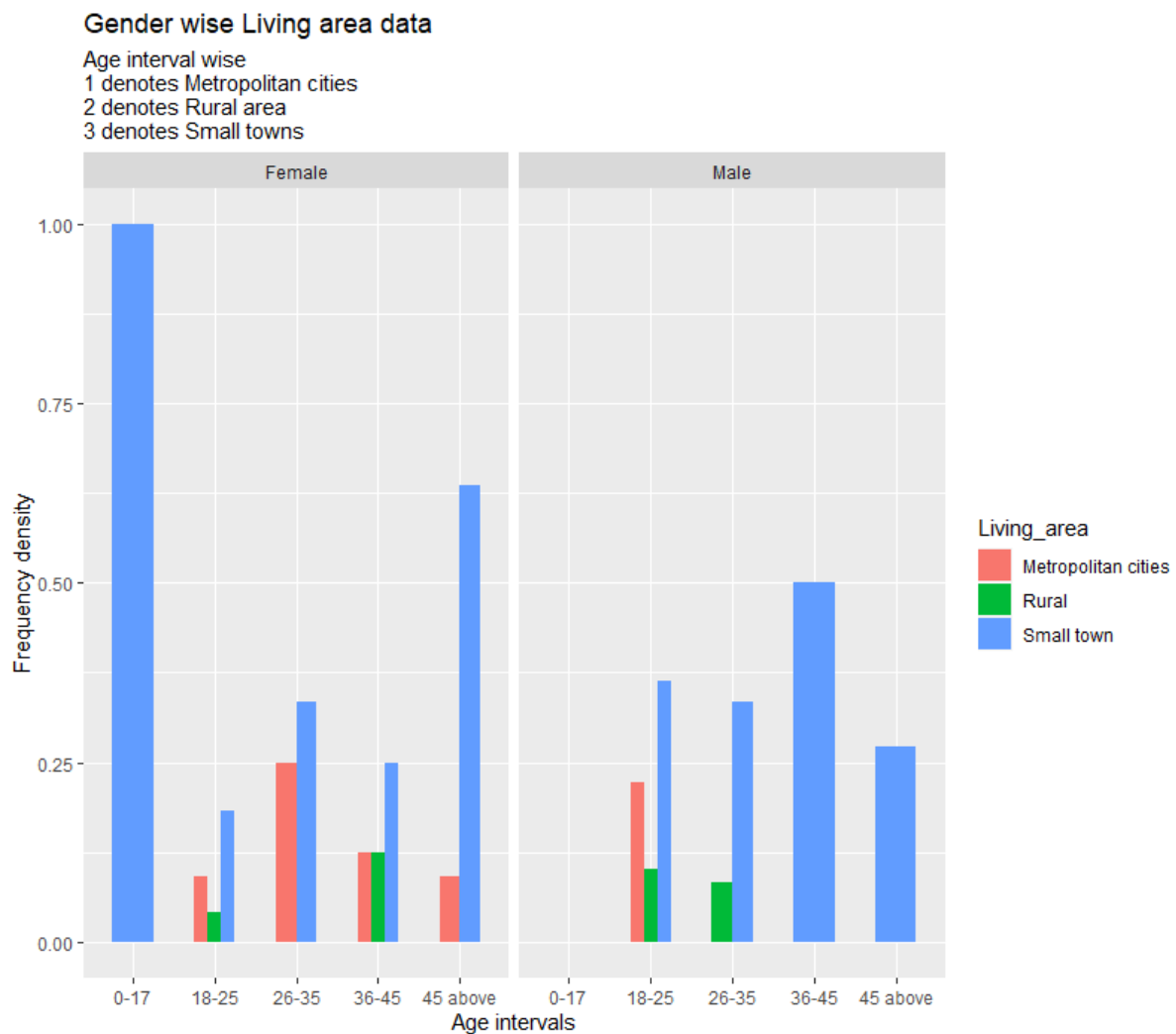


**Pie diagram to show Living area in the data**  
 '1' Represents Metroplitan area.  
 '2' Represents Rural area.  
 '3' Represents Small Town.

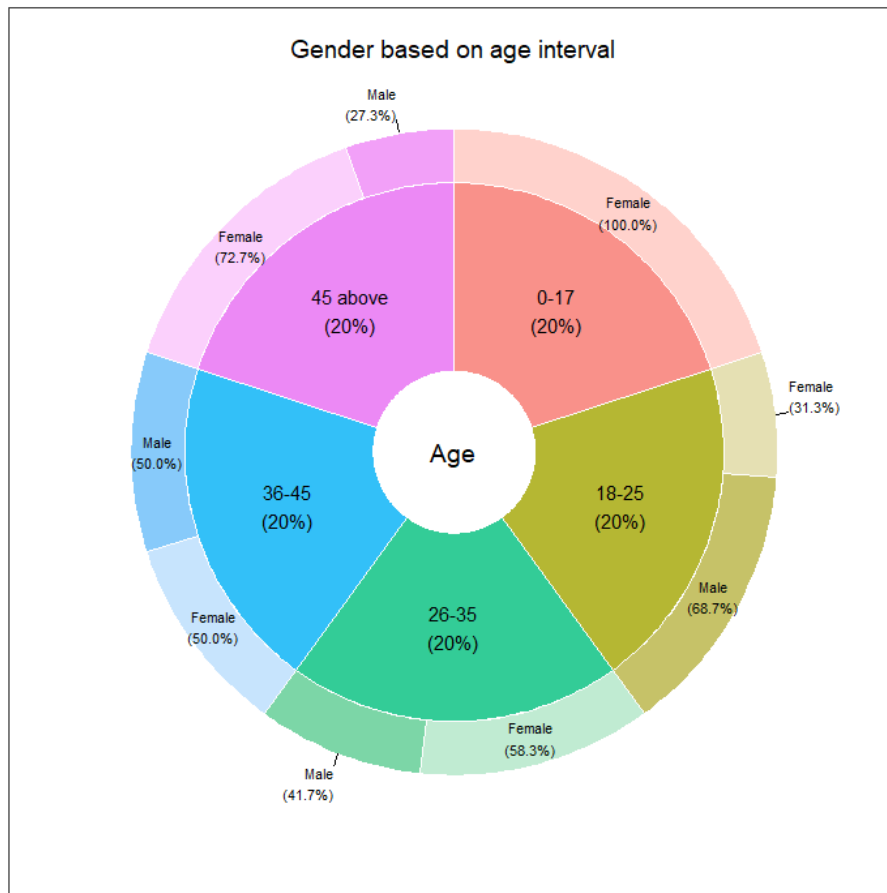


- Most of the respondents in our study come from small cities and minimum number of people are from rural area.
- Nearly 80 people are from small cities.
- Little more than 15 people are from rural area.
- Data is diverse enough and no group contains too less observations to give conclusive inferences.

Now, we will look into the composition more closely using the frequency density of these characteristics together,



- This is a frequency density plot of the number of people living in different areas with respect to male and female respondents.
- The age groups of 36-45 and 45 above are mostly from small towns for both male and female.
- Very few responses are from the rural area and the responses from the metropolitan cities are mainly in the 18-25 age group for the male population.
- The data could have been more diverse if we had more data from the rural areas.
- The responses in the age groups 36-45 and 45 above for the male population only came from small towns.
- The only response we got from the below 18 female respondent comes from small towns.



- In the age group below 18, we have got only one response which is female.
- In the age group 18-25, we have got near about 70 percent responses from the male subjects.
- In the age group 45 above, most of the respondents are female.
- In the age group 36-45 and 26-35 have nearly equal distribution of male and female respondents.
- The data would have been more conclusive and diverse if we had more data in the age group below 18.

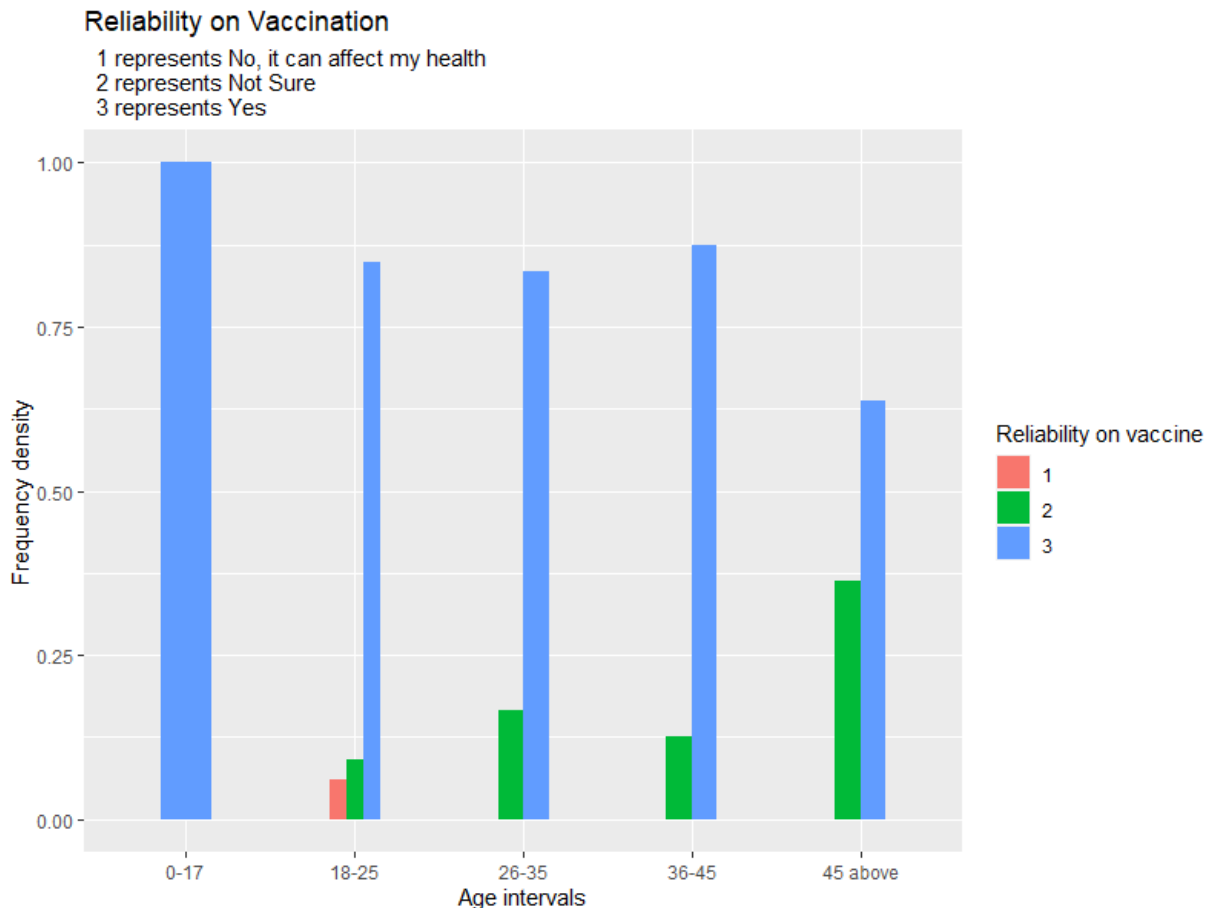
Now, we would dive into the exploratory data analysis (EDA) using the different columns in our data.

# DATA VISUALIZATION

## (EXPLORATORY DATA ANALYSIS)

We have already mentioned the relevant question in the abstract portion earlier. Now, we will take on those questions using visualization techniques from exploratory data analysis.

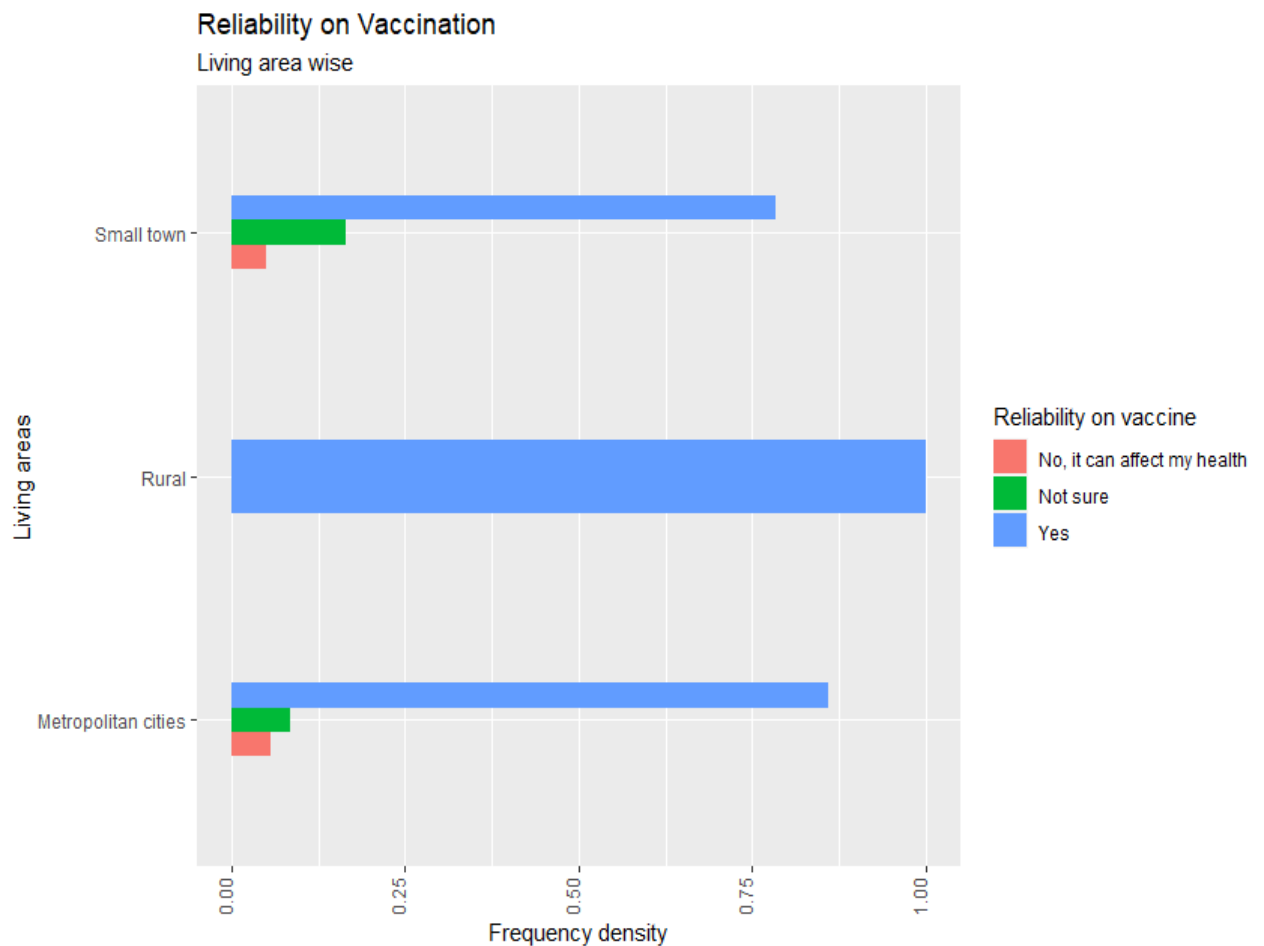
1.



- Interestingly, from the above multiple bar plot, it is clear that respondents above the age 45 are the ones who are most “Not sure” about the impact of the vaccine.
- The respondents in the age group 18-25 are most confident about the positive impact of the vaccine.
- From the data collected, it is also observed that the proportion of people who are not sure about the impact of the vaccine in the age group 26-35 are more than in the group 36-45, that is the respondents of age 26-35 are more sceptical than the people of age 36-45.

- The small proportion of respondents who believe that the vaccine can affect their health comes from the age group 18-25.

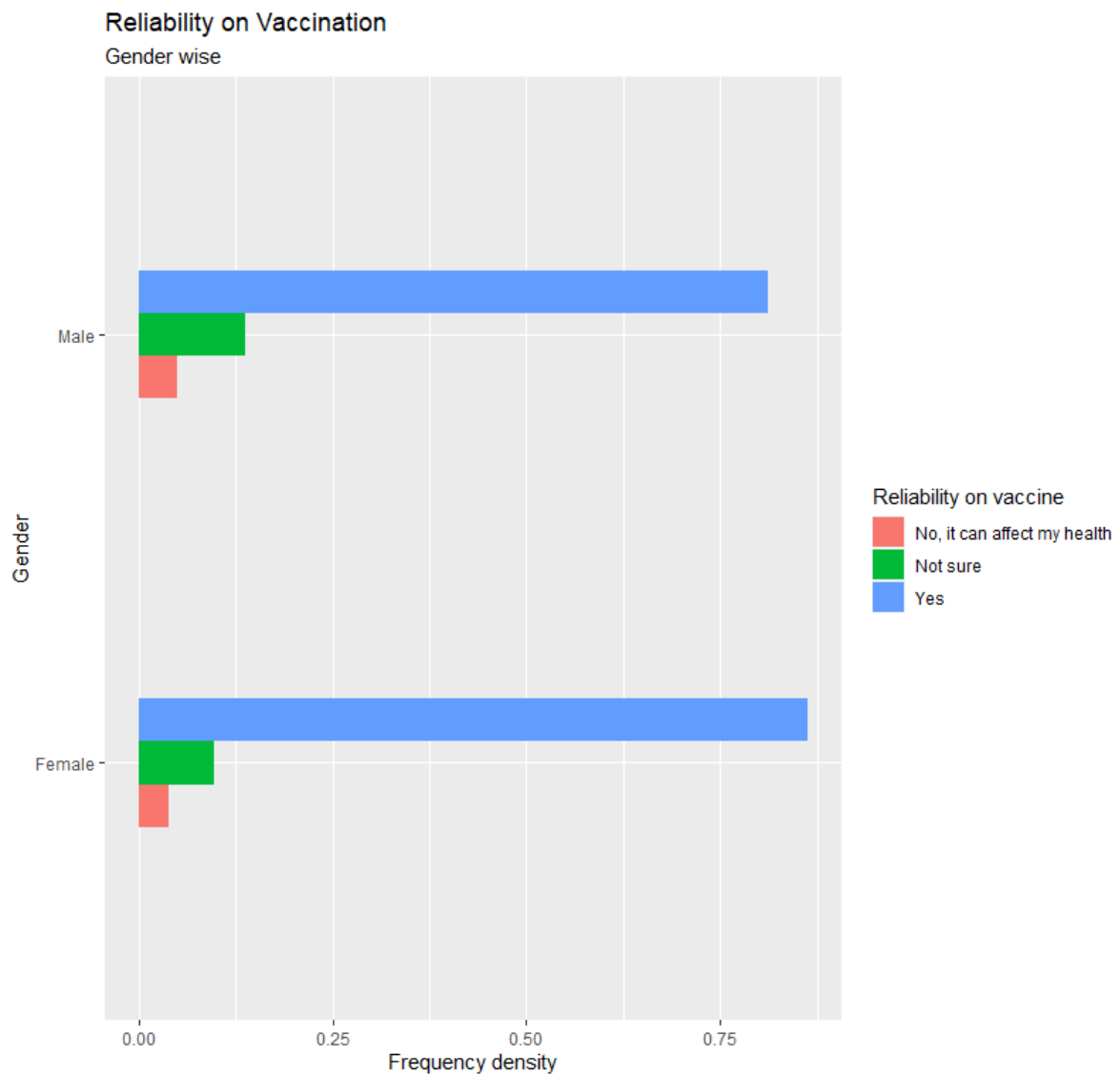
2.



The above bar plot also gives us some interesting observations-

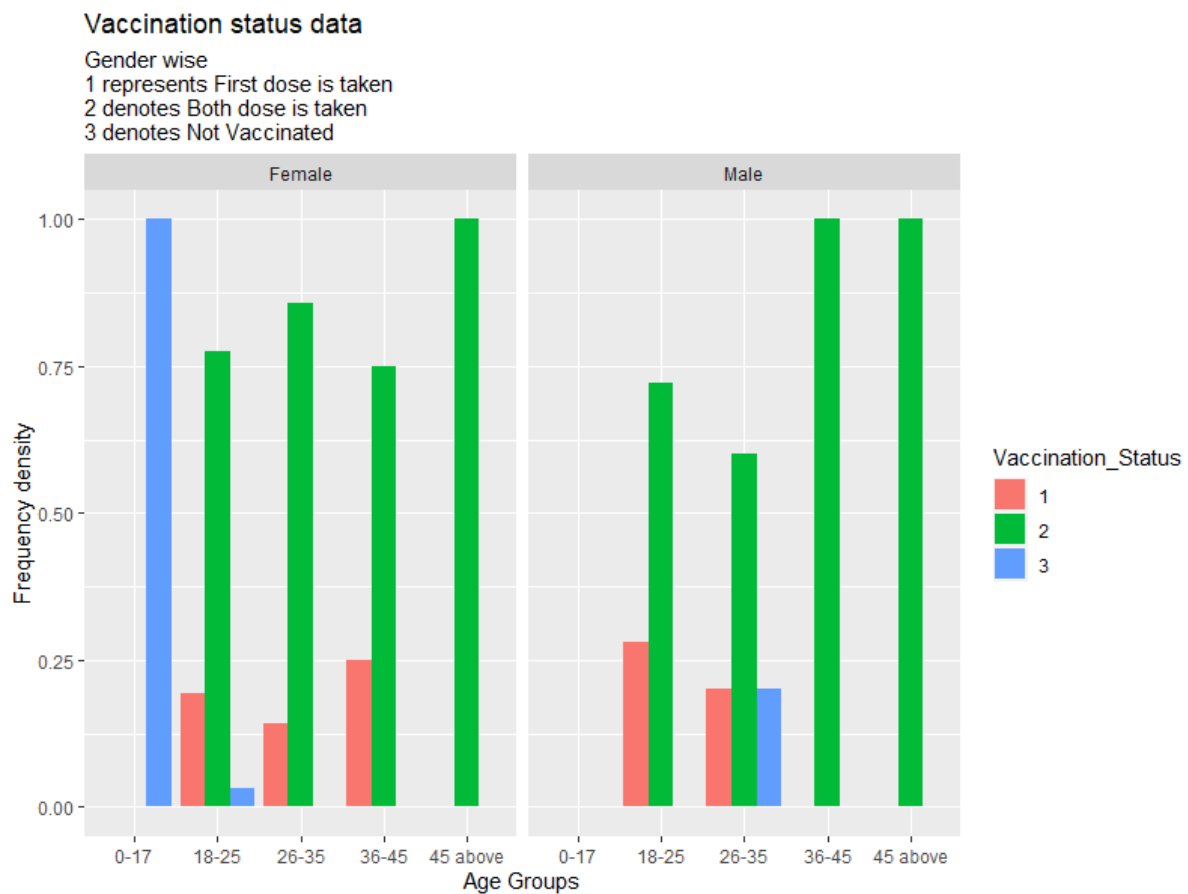
- All respondents from rural areas firmly believe in the positive impact of the vaccine.
- Whereas, maximum number of respondents who are not sure about the impact of the vaccine come from small towns.
- The proportion of people who believes in the negative impact of the vaccine comes from small town and metropolitan cities and the proportion in each group is near about equal.

3.



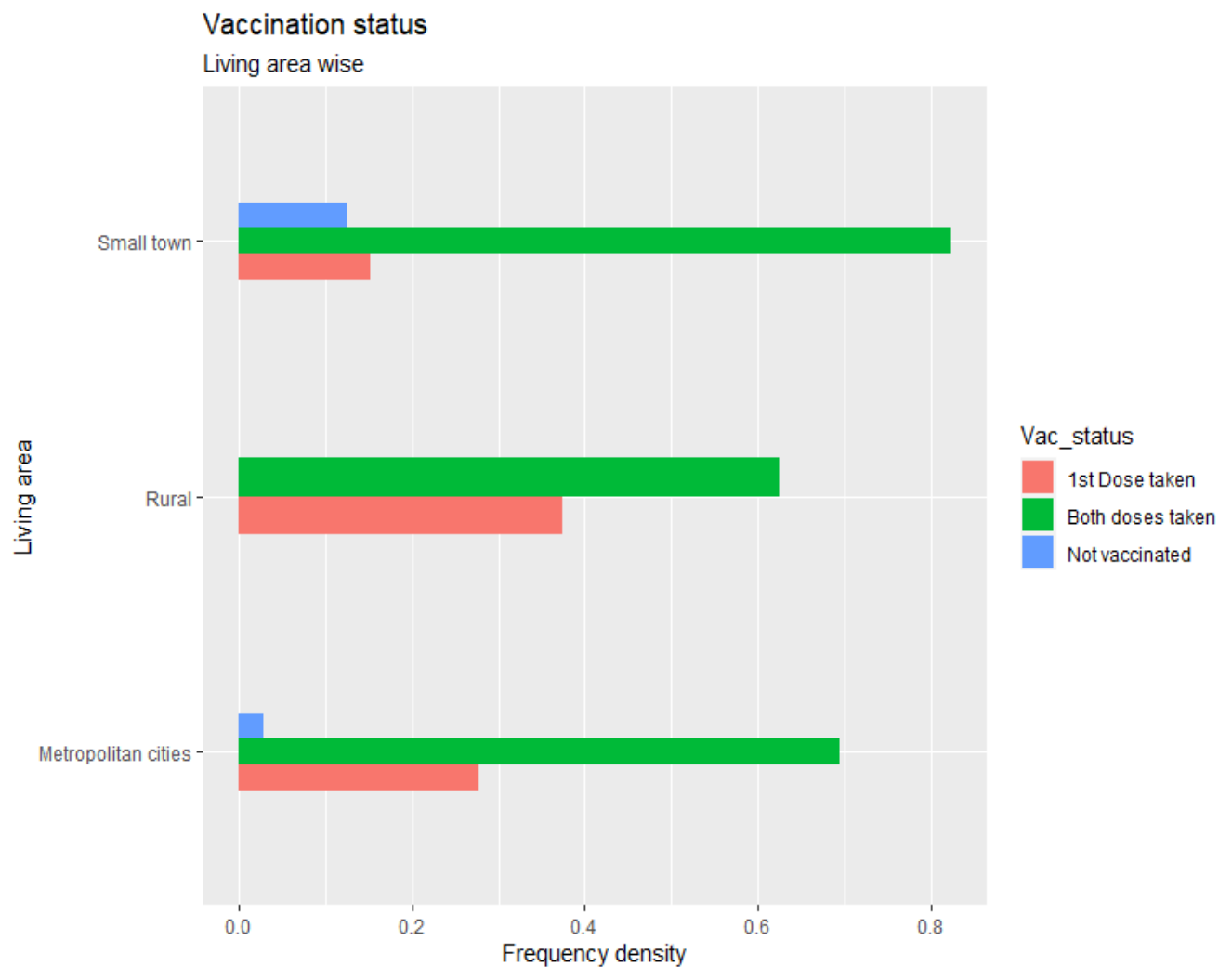
- The proportion of people who believe in positive impact of the vaccine is slightly more for the female respondents.
- Proportion of respondents from the male population are sceptical or not sure about the impact of the vaccine.
- Proportion of people who believe that vaccine can affect their health is more in case of male respondents.

4.



- For the age group 45 and above, in both male and female, all respondents are fully vaccinated.
- In the age group 36-45, all male respondents are fully vaccinated whereas in case of female 25 percent of them are partially vaccinated.
- Since vaccination for below 18 people are not initiated properly, that's why the only female responder of age below 18 is not vaccinated at all.
- For the group 18-25, the proportion of people who are partially vaccinated is more in case of male responders. Also, a small proportion of female responders in this group are not vaccinated at all.
- In age group 26-35, near about  $\frac{1}{4}$  th of the male population is not vaccinated at all and the proportion of partially vaccinated people is also greater than female responders.

5.

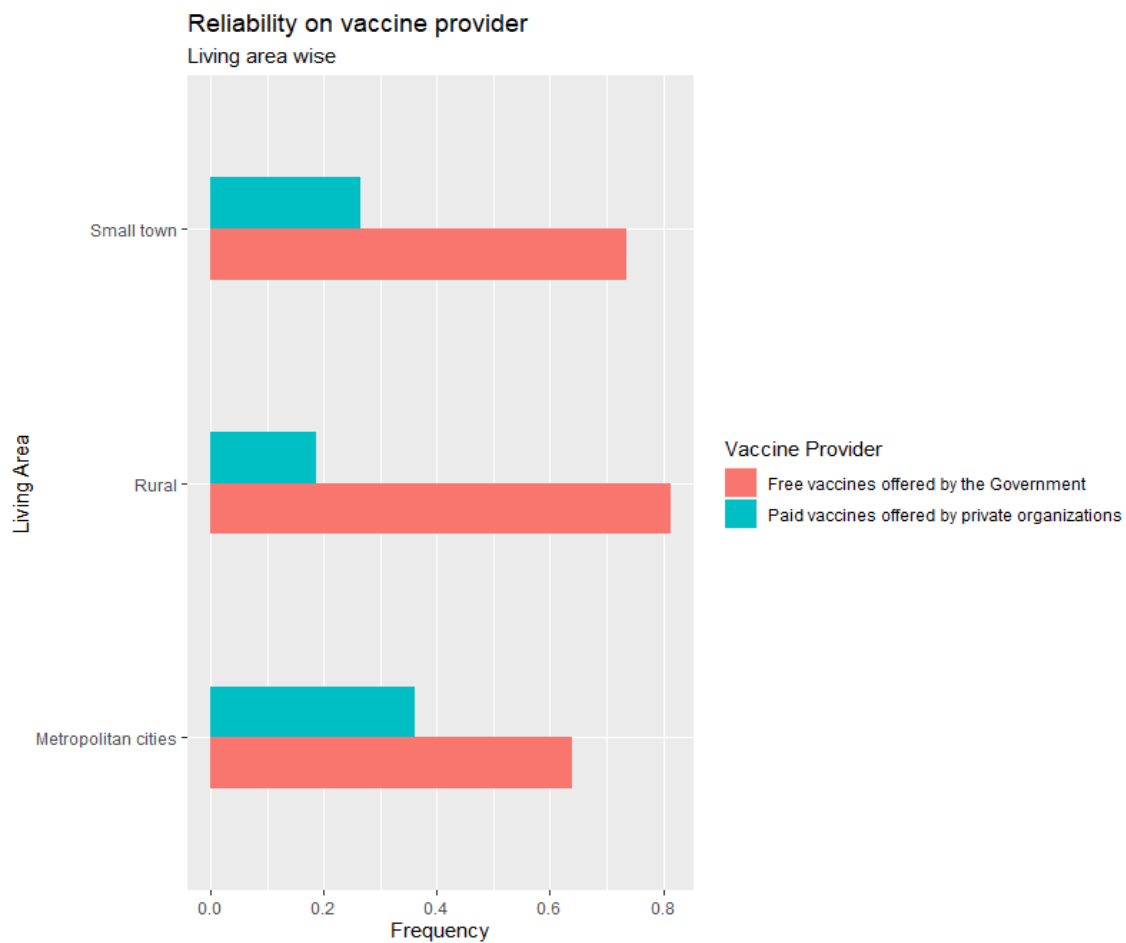


In the bar chart above, the vaccination status for the responder with respect to resident area type is represented. Here are some interesting inferences,

- It is clear that people from rural area are either fully vaccinated or partially vaccinated. Though the proportion of partially vaccinated people is highest in case of rural area.
- Proportion of people who are not yet vaccinated mostly come from small towns.
- Proportion of people who are fully vaccinated is maximum in case of small town.



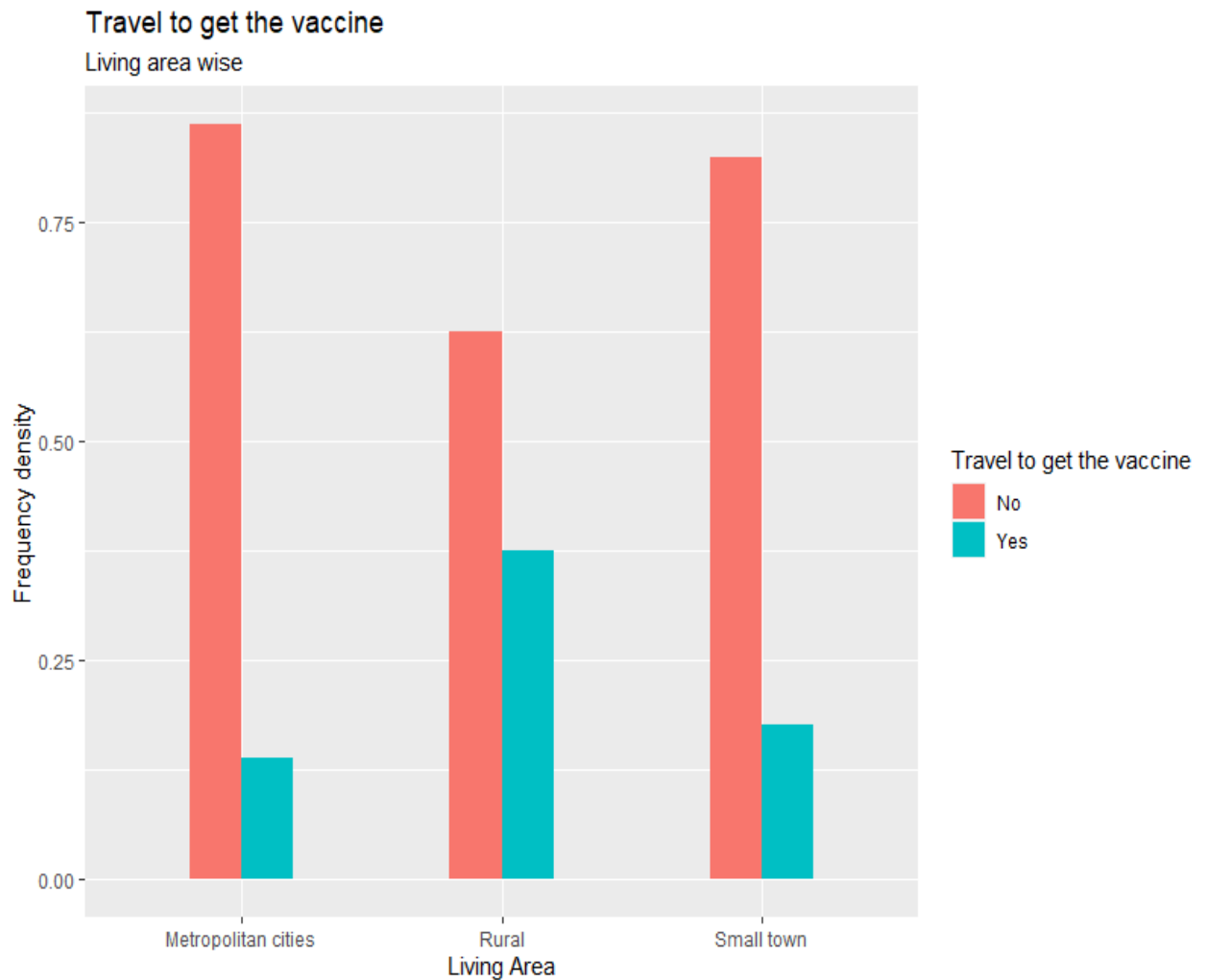
6.



The above bar plot represents the reliability on the vaccine providers for the responders categorised with respect to their resident area type. The relevant observations are,

- The proportion of people who are more inclined towards the free vaccine offered by the government is maximum in case of responders from the rural areas. These are the people who are least likely to get paid vaccines from private enterprises.
- The proportion of people who are more inclined towards the paid vaccine offered by the private bodies is maximum in case of responders from the metropolitan areas. Also, these are the people who show least trust on the free vaccines given by the Government.

7.

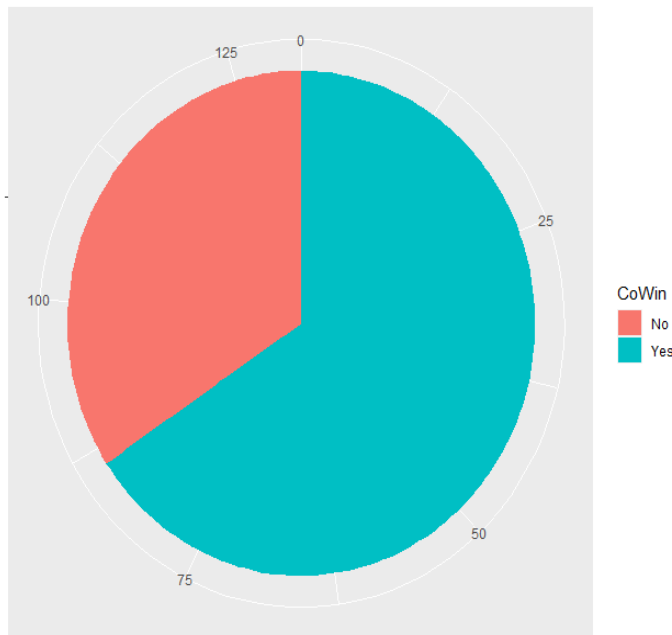


The multiple bar chart above represents the availability of the vaccines in the different types of resident area of the responders. We can make the following observations,

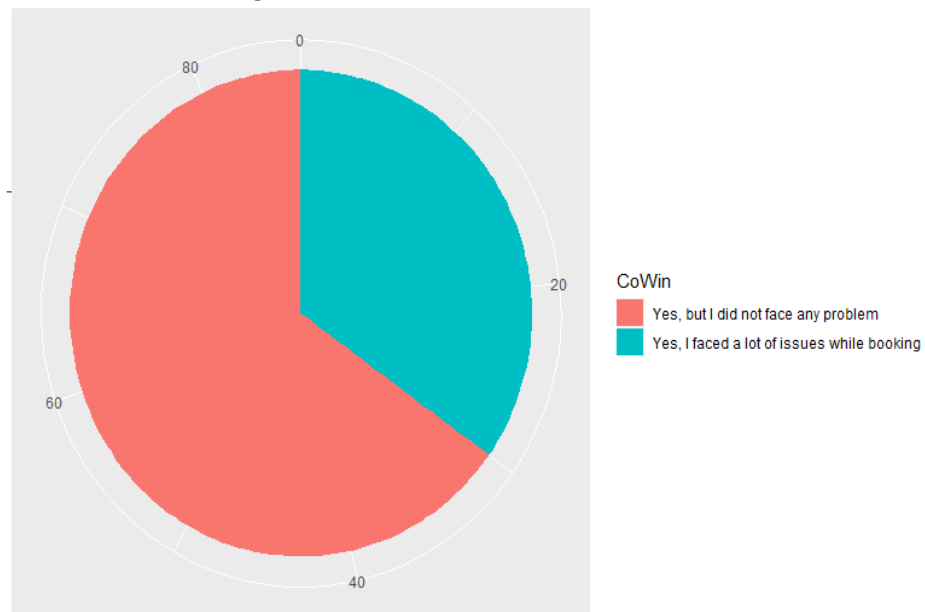
- The proportion of people who had to travel to get the jab is highest for the responders from the rural areas and least for the responders from the metropolitan areas.
- From the chart above we can say that, vaccines were least available in the rural areas, followed by small towns and vaccines were most available in the metropolitan areas.

8.

Percentage of Cowin usage

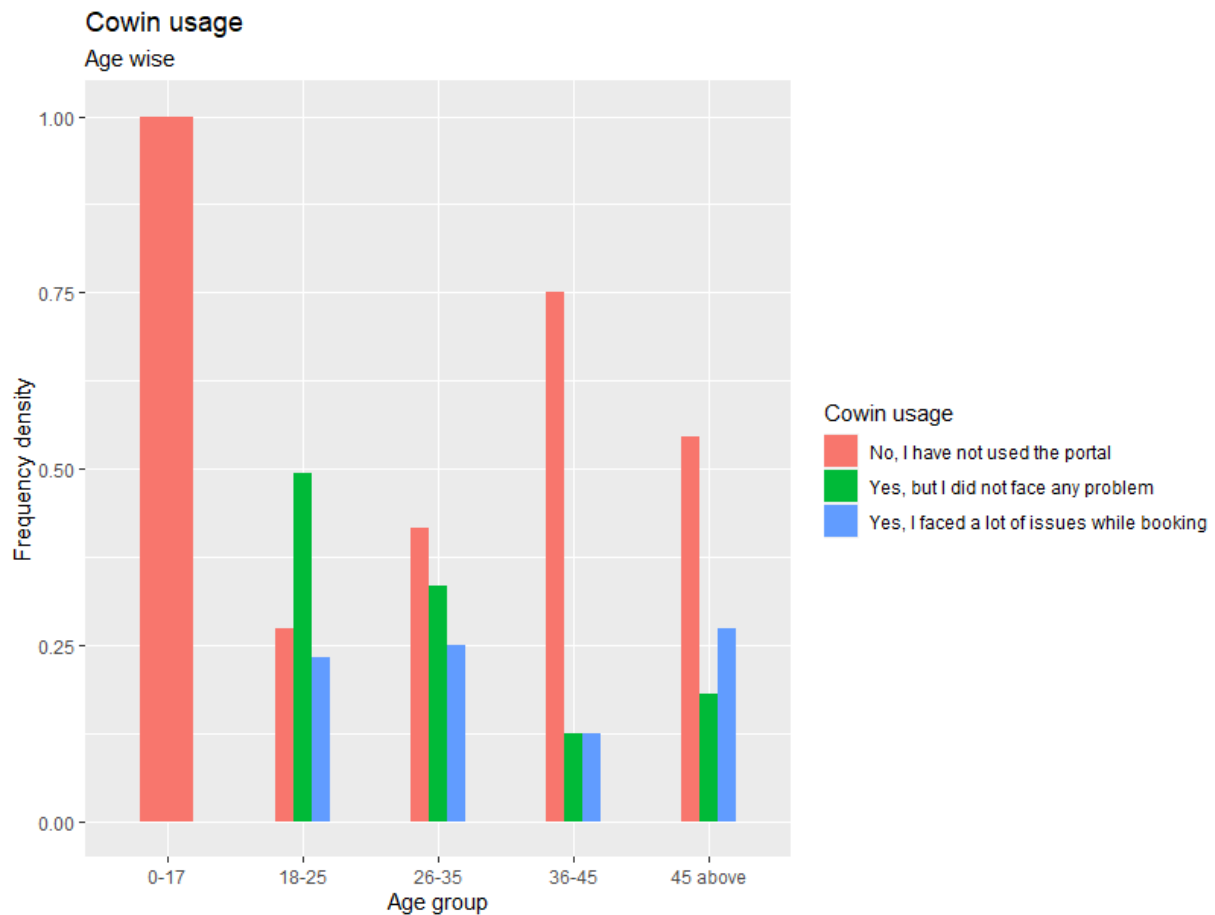


Problem faced while using Cowin



- In our data set, the number of people who has used the CoWin portal is nearly 65 percent while the rest of the people didn't use the CoWin portal.
- Out of the 65 percent of people who used the portal, 40 percent of the people faced no problem while using the portal and the rest 60 percent faced a lot of issues while using the portal.

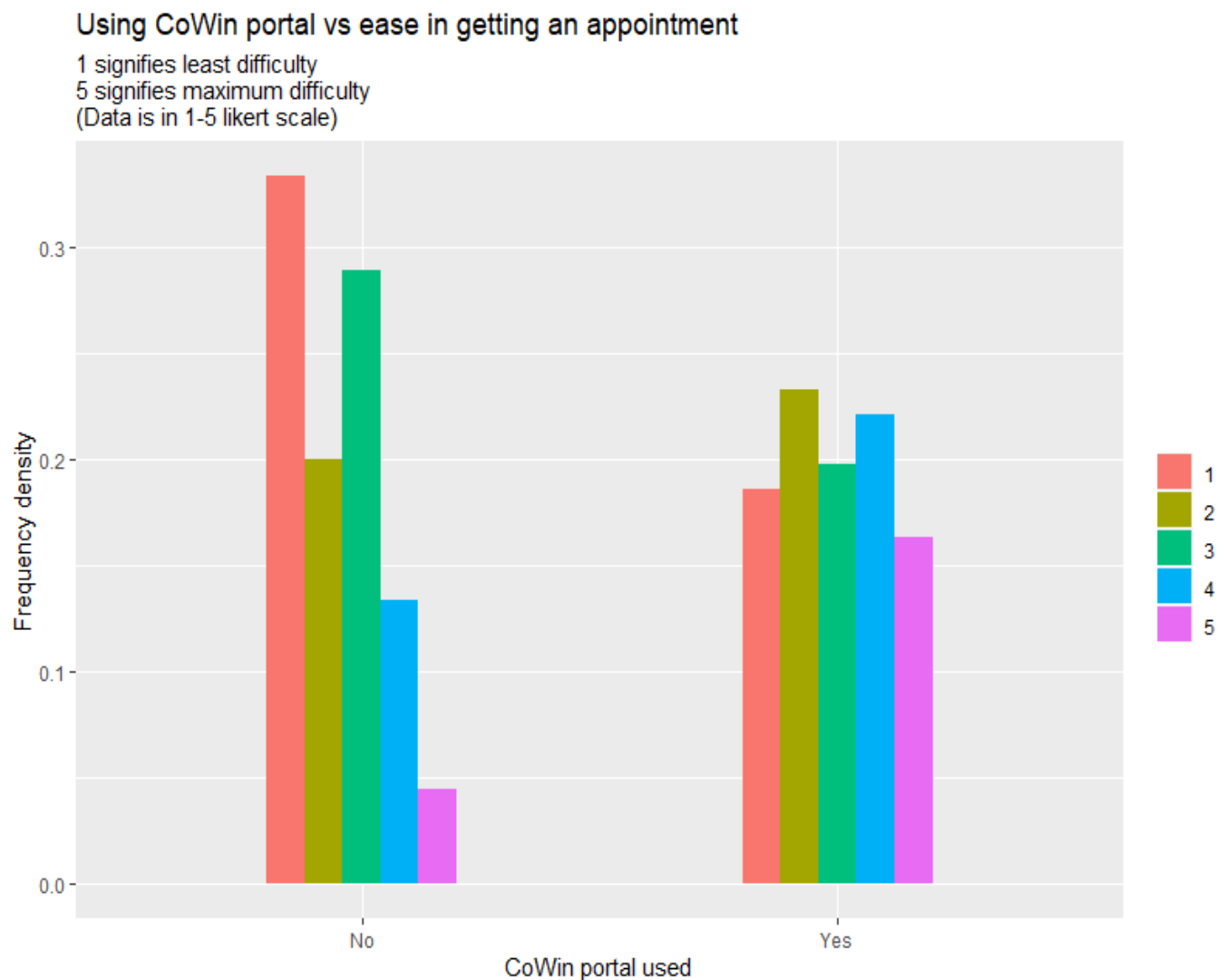
9.



The above multiple bar plot tells us whether a responder from a specific age group has used the CoWin portal or not for booking appointment for the vaccination and faced any difficulties or not. The observations are following,

- The proportion of people who haven't used the portal is maximum in case of the age group 36-45, followed by the age group 45 above.
- The proportion of people who used the portal for booking an appointment for vaccination and faced a lot of issues is highest for the people of age 45 and above, followed by the people in age group 26-35 and 18-25 and it is least in the age group 36-45.
- The proportion of people who did use the portal and the booking process was fluent is maximum for the age group 18-25 and least for the age group 36-45, followed by the age group 45 and above.

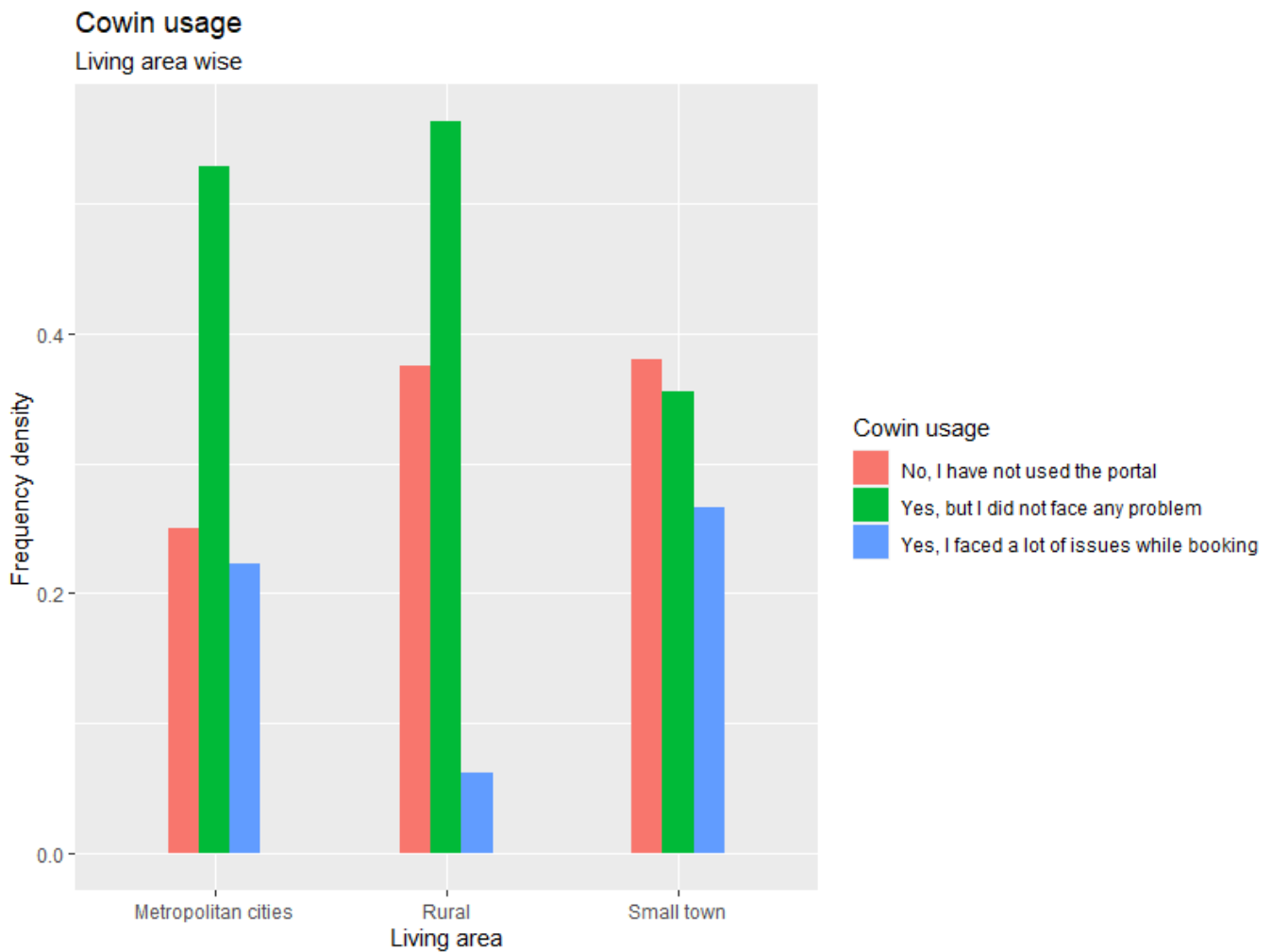
10.



This multiple bar plot tells us the relationship between using of CoWin portal and ease in booking a slot. The observations are,

- Those proportion of people who had face the maximum difficulty, it is very likely that they had had used the CoWin portal for getting the appointment.
- If an individual hadn't used the portal to book the slot, there is a 35 percent chance that he has faced least difficulty, whereas the same percentage is near about 20 percent in case of people who had used the portal.
- Interestingly enough, though the proportion of people with least difficulty is higher in case of not using the portal, proportion of people who have faced the average level of difficulty is way higher in case of the former group.

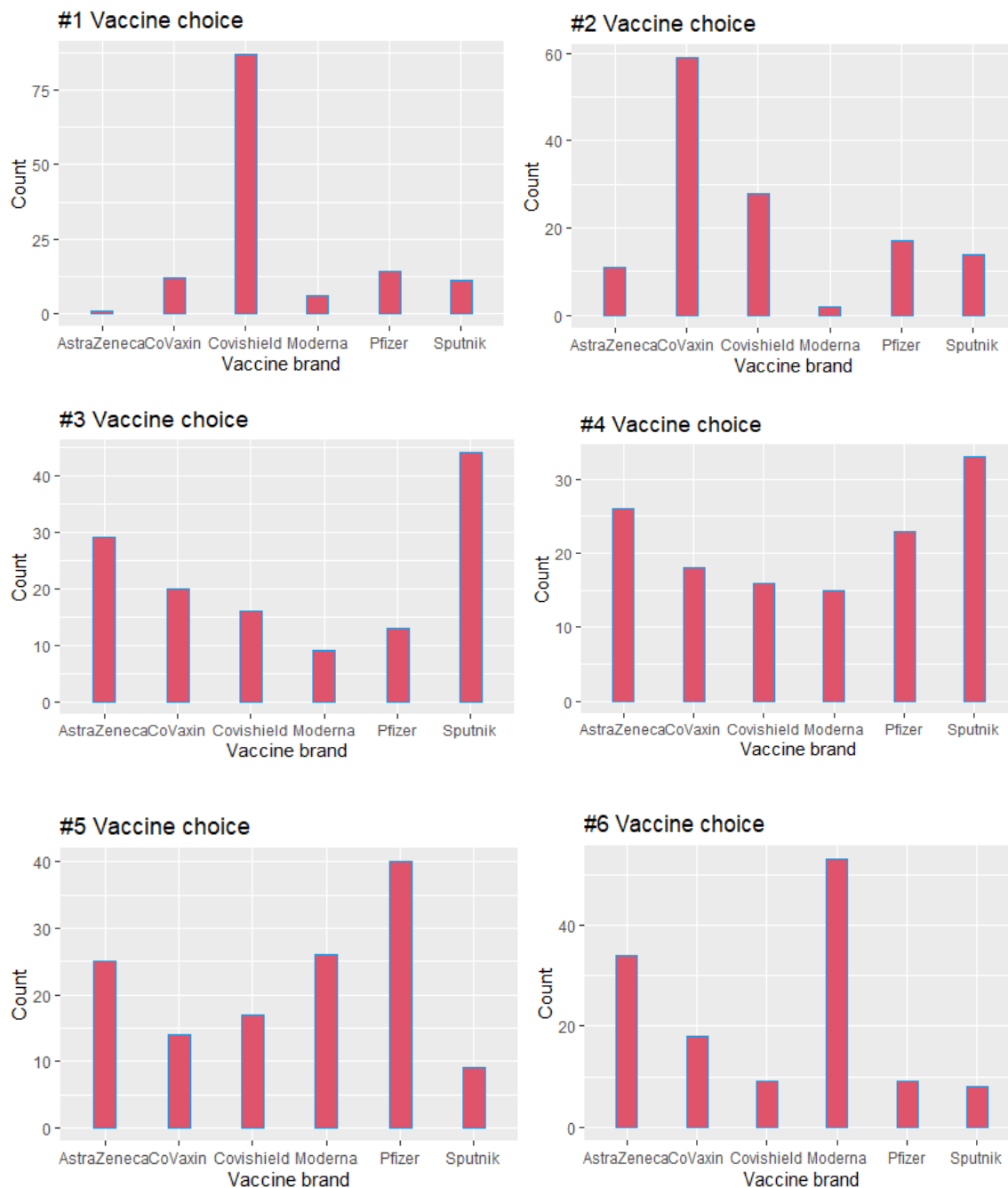
11.



The above multiple bar plot tells us whether a responder from a specific type resident area has used the CoWin portal or not for booking appointment for the vaccination and faced any difficulties or not. The observations are following,

- The proportion of people who haven't used the portal for booking an appointment is lowest for responders from metropolitan cities. For rural and small town, this proportion is almost same.
- The proportion of people who could book a slot through the portal fluently is highest for rural areas followed by small towners and it is least for the people from metropolitan cities.
- The proportion of people who had a hard time booking a slot through a portal is highest for small town and it is least for rural areas.

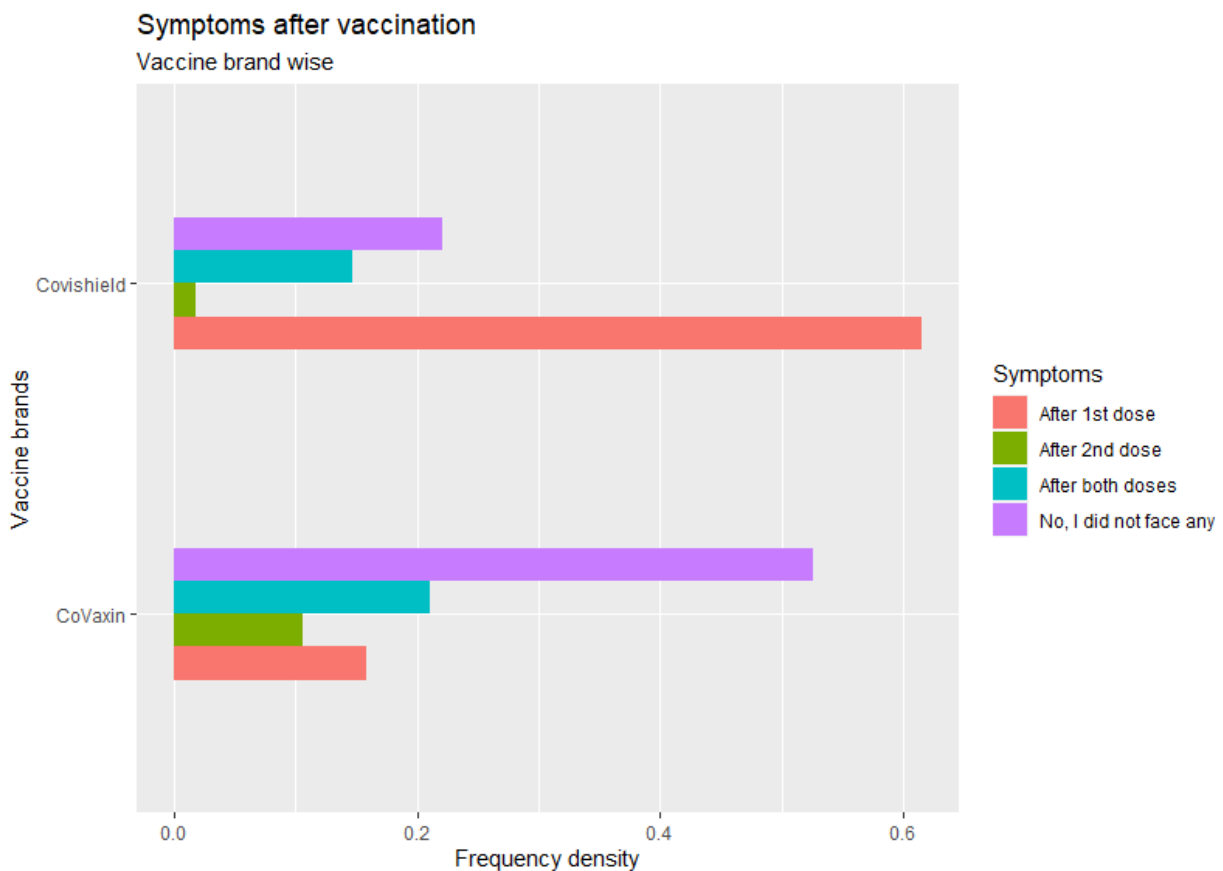
12.



The six bar plots above are extracted from the six columns which gives us the choices of vaccines preferred by the responders. We can make the following observations-

- As first choice of vaccine the most preferred brand is Covishield.
- As second choice of vaccine the most preferred brand is Covaxin.
- As fifth choice of vaccine the most preferred brand is Pfizer.
- As last choice of vaccine the most preferred brand is Moderna.

13.

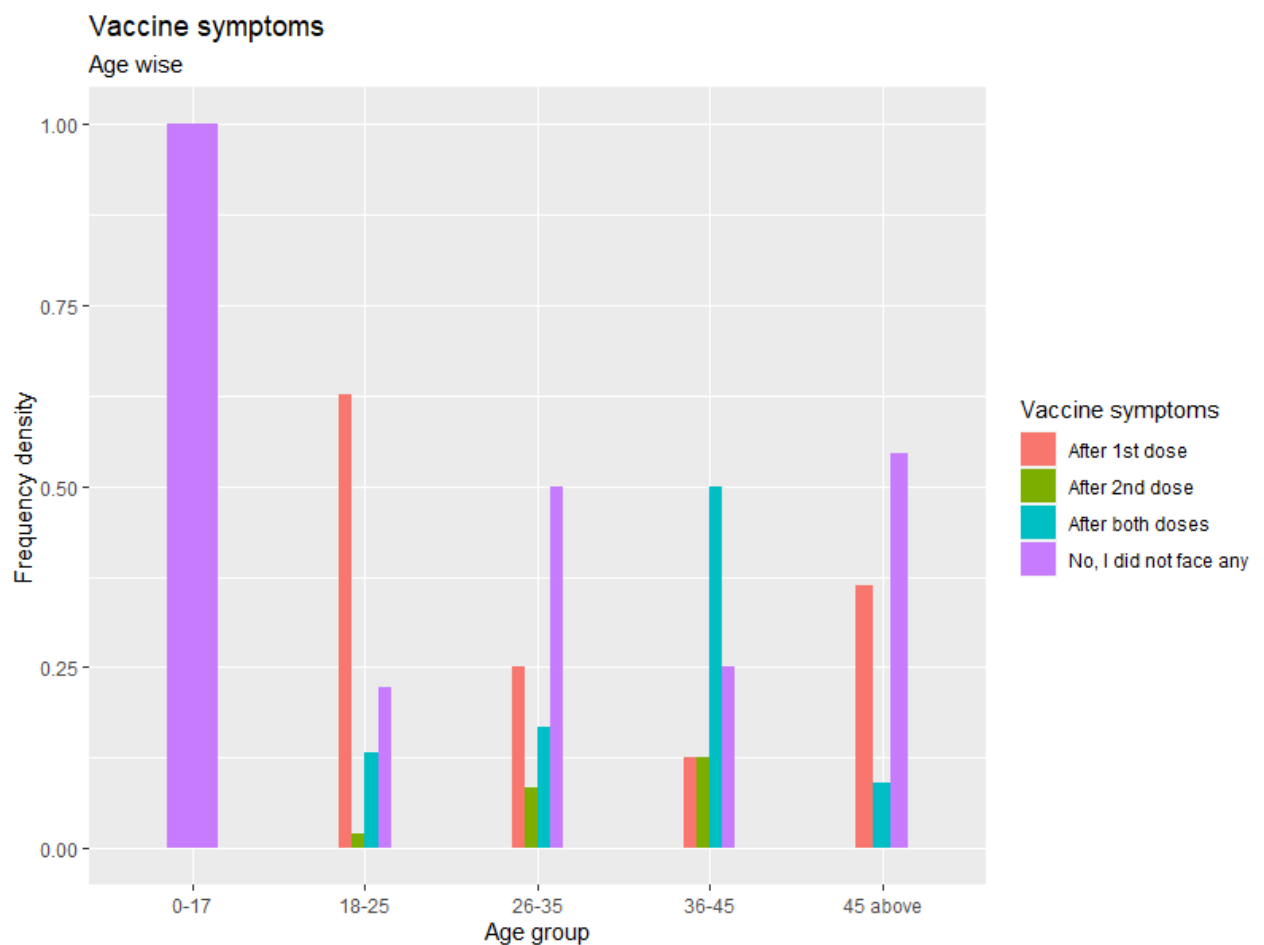


The above bar chart between vaccine brands and post vaccination symptoms tells us the following things,

- The proportion of people who had suffered from post vaccination symptoms like fever, cold etc after 1<sup>st</sup> dose is very high (slightly more than 60 percent) in case of Covishield receivers than the people who got Covaxin (slightly more than 15 percent).
- The proportion of people who had suffered from post vaccination symptoms after 2<sup>nd</sup> dose is very high (10 percent) in case of Covaxin receivers than the people who got Covishield (slightly less than 5 percent).
- The proportion of people who had suffered from post vaccination symptoms after both doses is near about same for both Covaxin and Covishield receivers.
- Among the Covaxin receivers, the proportion of people who didn't suffer from any post vaccination symptoms is very high (slightly more than 50 percent) compared to Covishield receivers (slightly more than 20 percent).



14.

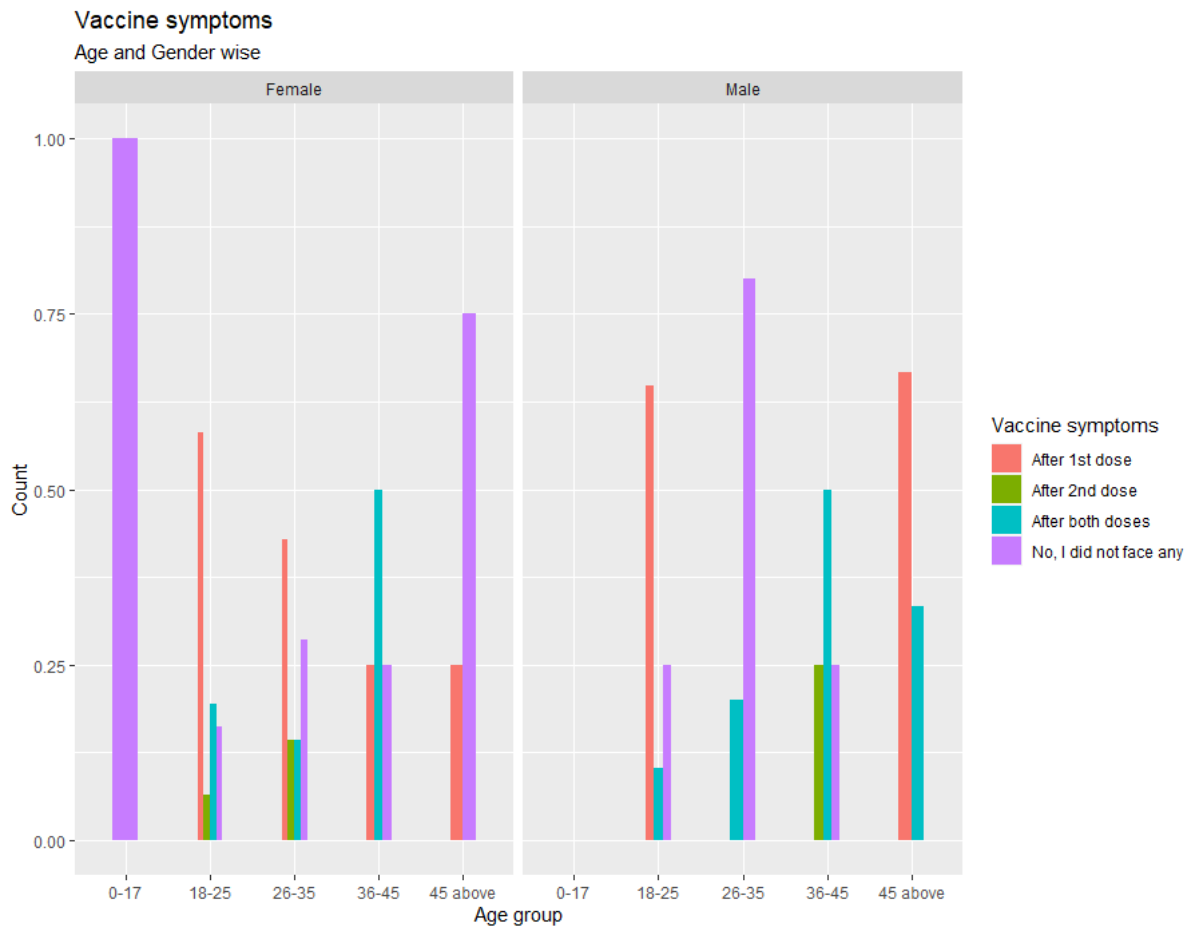


The above bar plot shows us the post vaccination symptoms categorized by age groups.

The important observations are-

- The responders in the age group 18-25 shows the maximum proportion of people who had suffered from post vaccination symptoms after 1<sup>st</sup> dose, whereas responders in the age group 36-45 shows the least.
- People in the age groups 26-35 and 45 above shows the maximum proportion of people who did not suffer from any post vaccination symptoms.
- There are very few people in the data set who suffered after the 2<sup>nd</sup> dose. Among them who did suffer, the age group 36-45 has the highest proportion in that.
- People in the age group 36-45 have the highest proportion of suffering after both the doses.

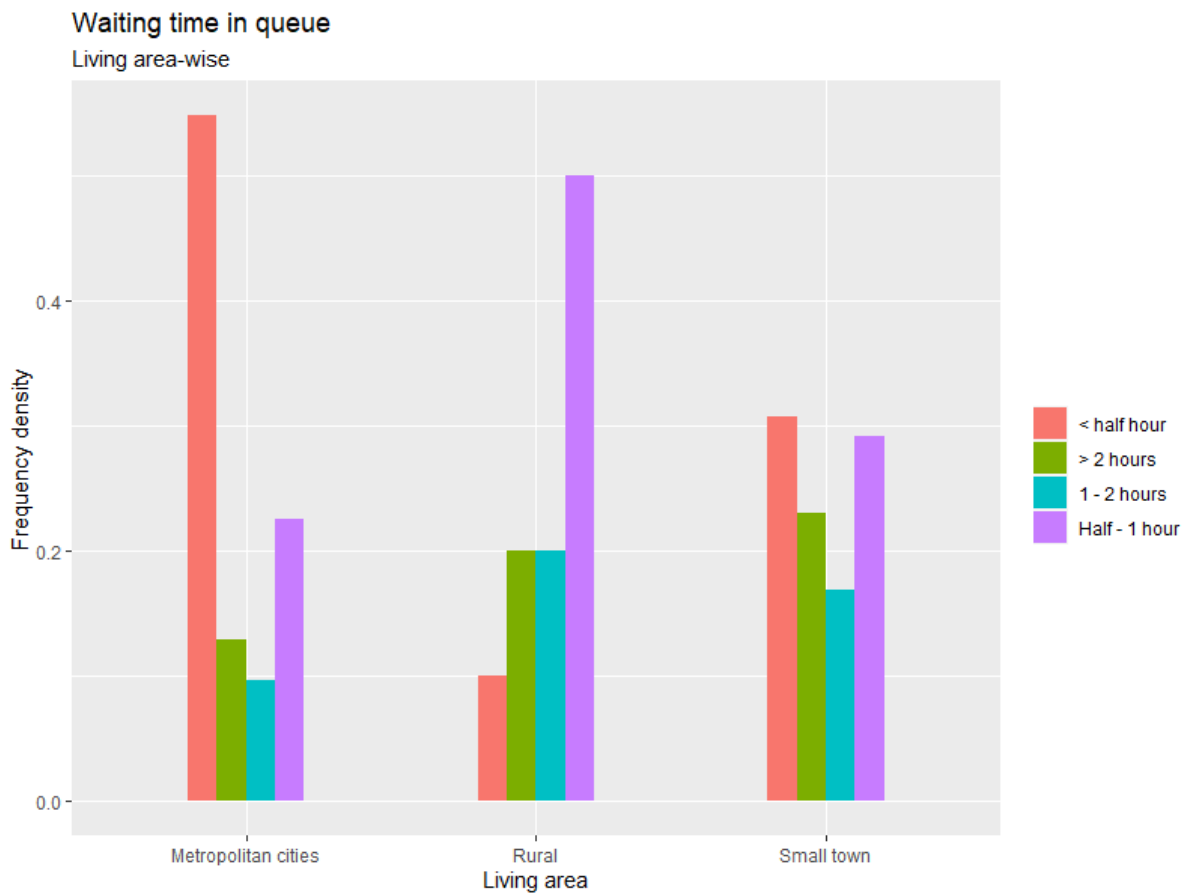
15.



The above diagram shows us the nature of post vaccination symptoms across different age groups categorized by gender of an individual. The observations that we can make are,

- In the age group 18-25, male responders are more prone to suffer from post vaccination symptoms after the 1<sup>st</sup> dose and also more prone to not suffer from any symptoms.
- For the age group 45 above, male subjects are way more likely to suffer from symptoms after the 1<sup>st</sup> and after both doses than females.
- In the age group 36-45, male responders are suffering from symptoms after the 2<sup>nd</sup> dose, whereas the female responders are suffering mainly after the 1<sup>st</sup> dose.
- In the age group of 26-35, male responders are mostly not suffering from any symptoms, whereas almost 50 percent of the females of the same age group are suffering after 1<sup>st</sup> dose.

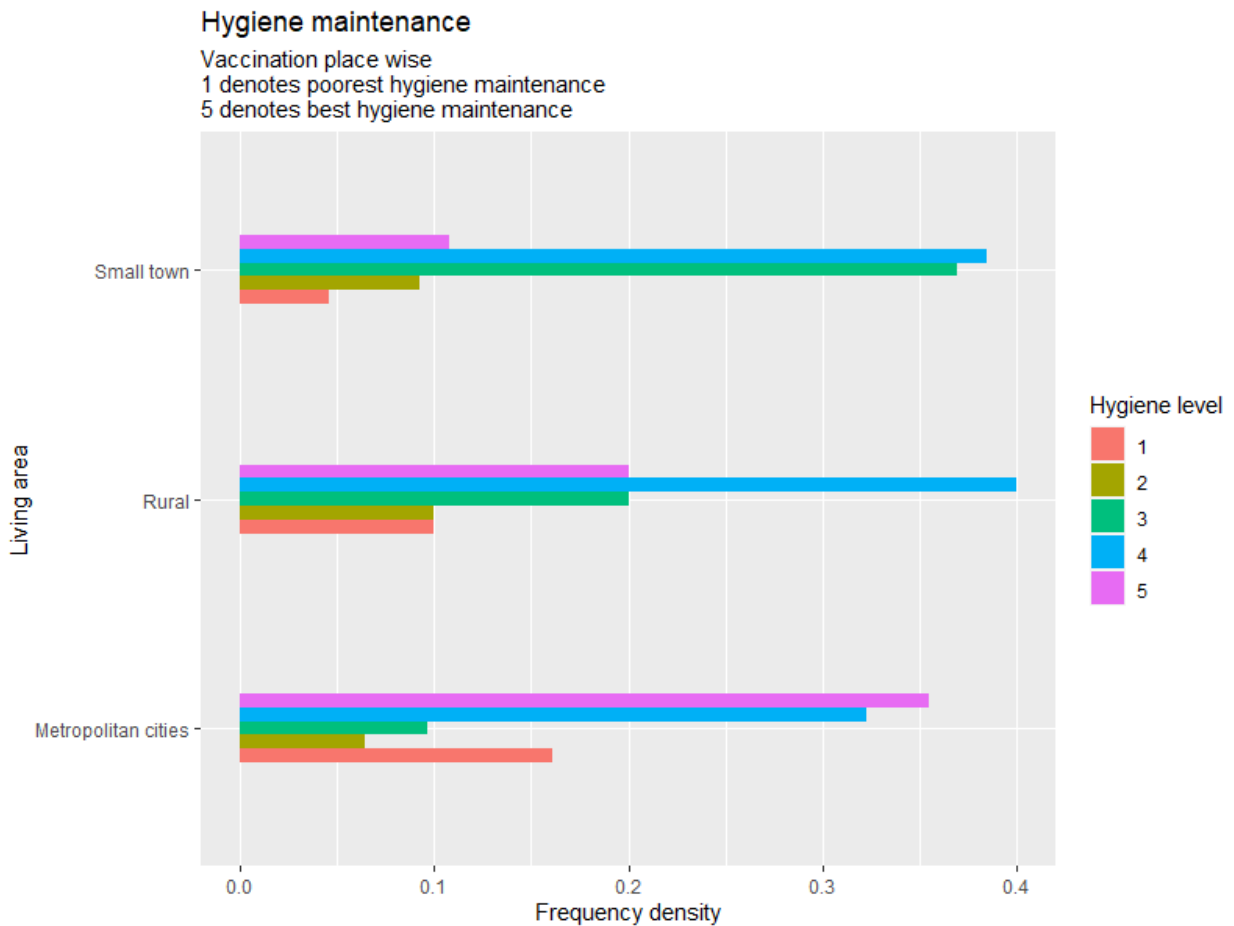
16.



The above chart represents the waiting time in the queue at the vaccination centre for an individual who was vaccinated in his or her own resident area. The relevant observations are,

- The proportion of people whose waiting time is less than half hour is mostly resident of a certain metropolitan city.
- Residents of rural areas shows the maximum proportion of people whose waiting time was half to one hour in the vaccination centre.
- Residents of rural areas shows the maximum proportion of people whose waiting time was one to two hours in the vaccination centre.
- The proportion of people whose waiting time is more than two hours is highest for the residents of small towns.
- From these observations, we can say that the experience regarding waiting time in queue at the vaccination centres has been best for the people in metropolitan cities followed by rural areas.

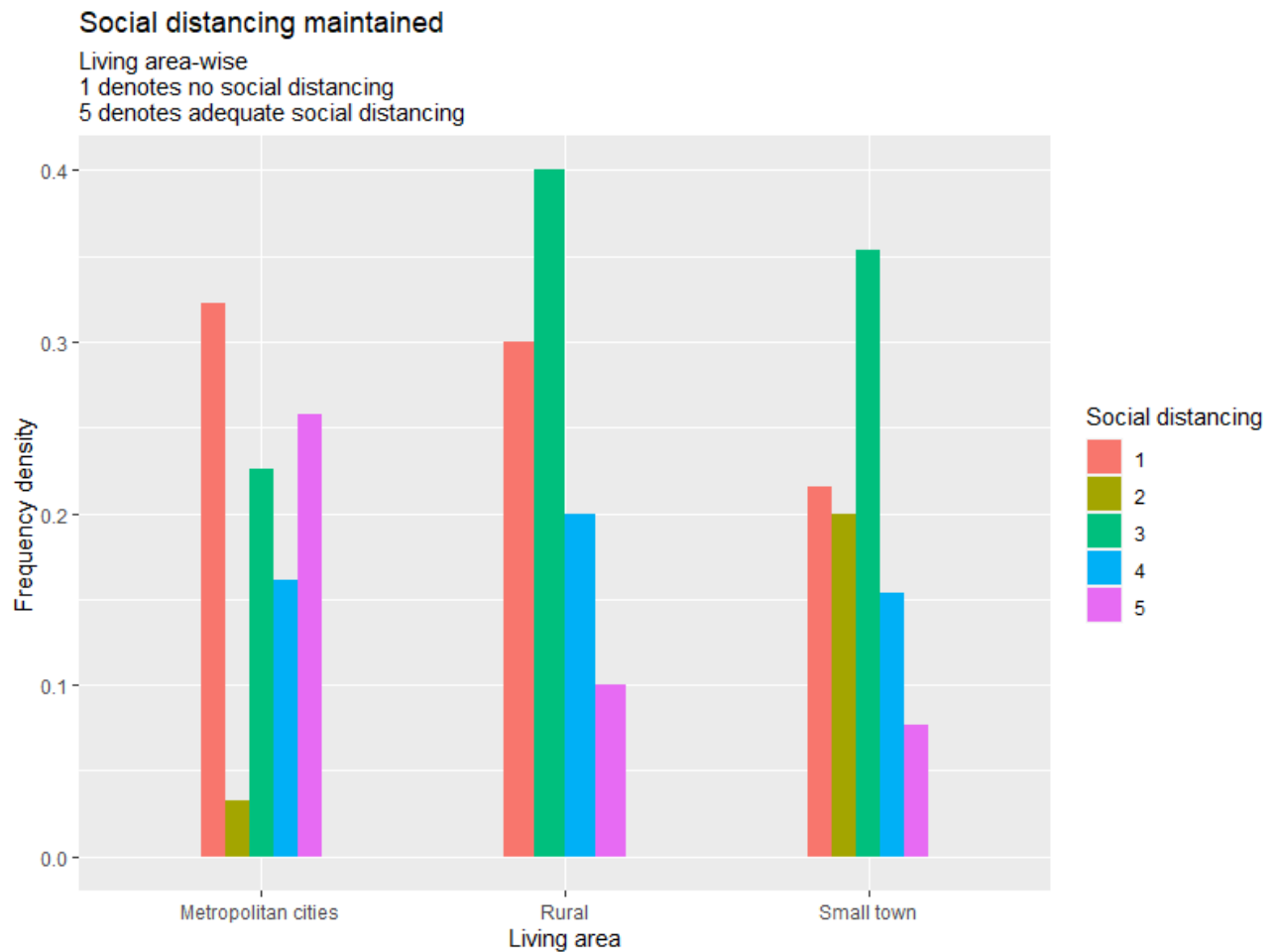
17.



The above chart represents the feedback on the hygiene maintained at the vaccination centre by an individual who was vaccinated in his or her own resident area. The relevant observations are,

- Metropolitan cities show the maximum number of people who experienced the best satisfactory level of hygiene maintained at the vaccination centre. Though this proportion is way less than the proportion of people living in metropolitan cities who witnessed the worst satisfactory level of hygiene and it is least for rural areas.
- The proportion of people who experienced average satisfactory level of hygiene maintained at the vaccination centre is highest for small town.
- Since, a greater number of people facing worst hygiene issues in metropolitan cities, so we cannot say that the hygiene experience is best here. Then we can say people living in rural area are most satisfied with the hygiene maintenance in their vaccination centres.

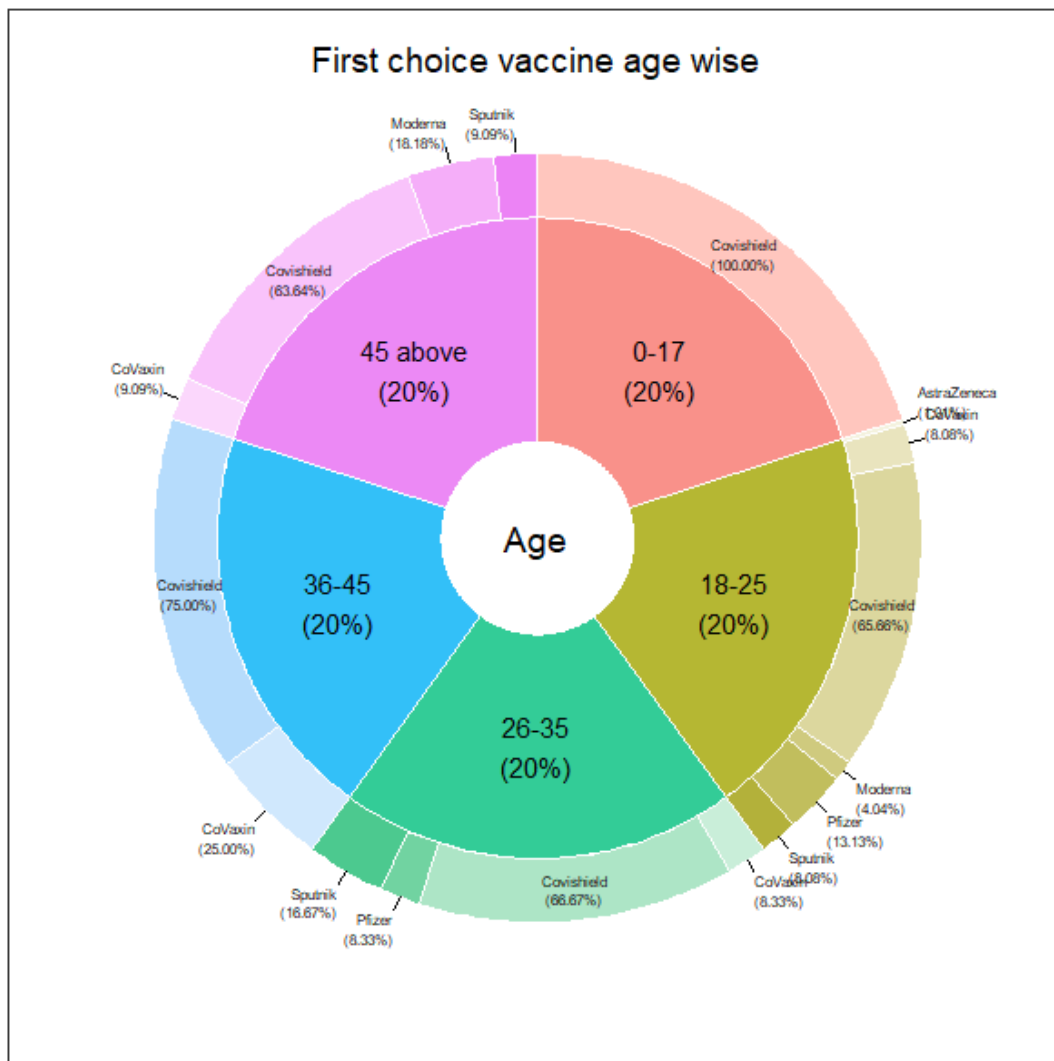
18.



The above chart represents the feedback on the social distancing maintained at the vaccination centre by an individual who was vaccinated in his or her own resident area. The relevant observations are,

- Metropolitan cities show the maximum number of people who experienced the best satisfactory level of social distancing maintained at the vaccination centre. Though it also shows maximum proportion of feedbacks of worst satisfactory level of social distancing maintained at the vaccination centre.
- The proportion of people who experienced average satisfactory level of social distancing maintained at the vaccination centre is highest for rural areas.
- Since, a greater number of people facing worst satisfactory social distancing issues in metropolitan cities, so we cannot say that the social distancing practised is best there. Then we can say people living in rural area are most satisfied with the social distancing practised in their vaccination centres.

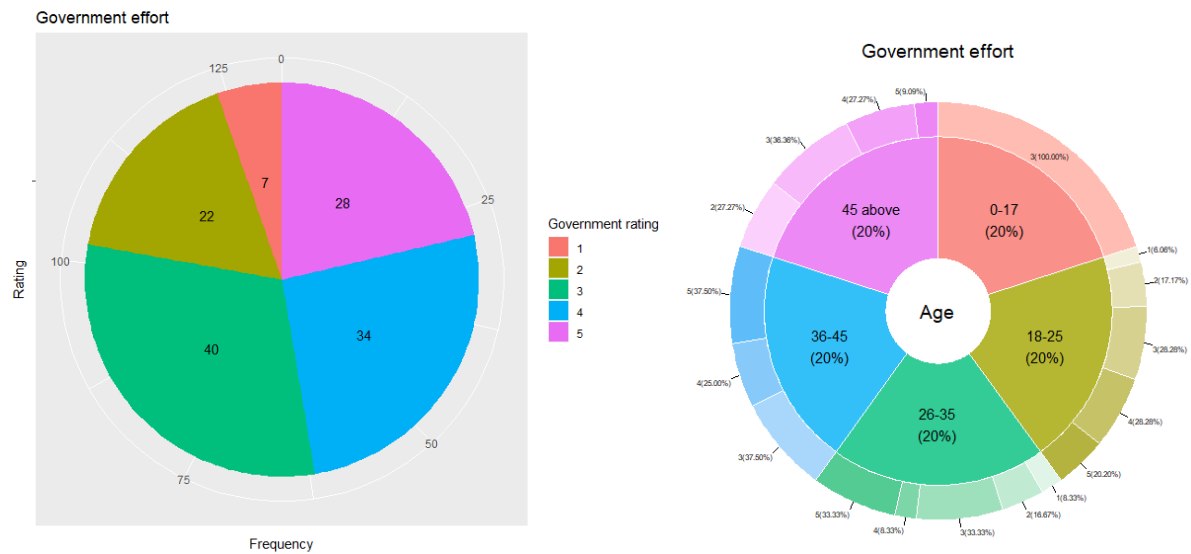
19.



The above pie donut plot is the representation of first choice vaccine with respect to the age group of the respondents. The relevant observations are,

- Through out all the age groups, the most preferred first choice vaccine is Covishield as can be seen from the above diagram.
- In the age group 18-25, the second most preferred of first choice vaccine brand is Pfizer. For the 26-35 age group, Sputnik seems to be the second most preferred of first choice vaccine brand. For the 36-45 age group, Covaxin is the second most preferred while for 45 and above Moderna seems to be the one.

20.



The above plots represent the feedback given by the people about the Govt.'s effort in conducting the process of vaccination by supplying enough vaccines for the people and making a safe environment.

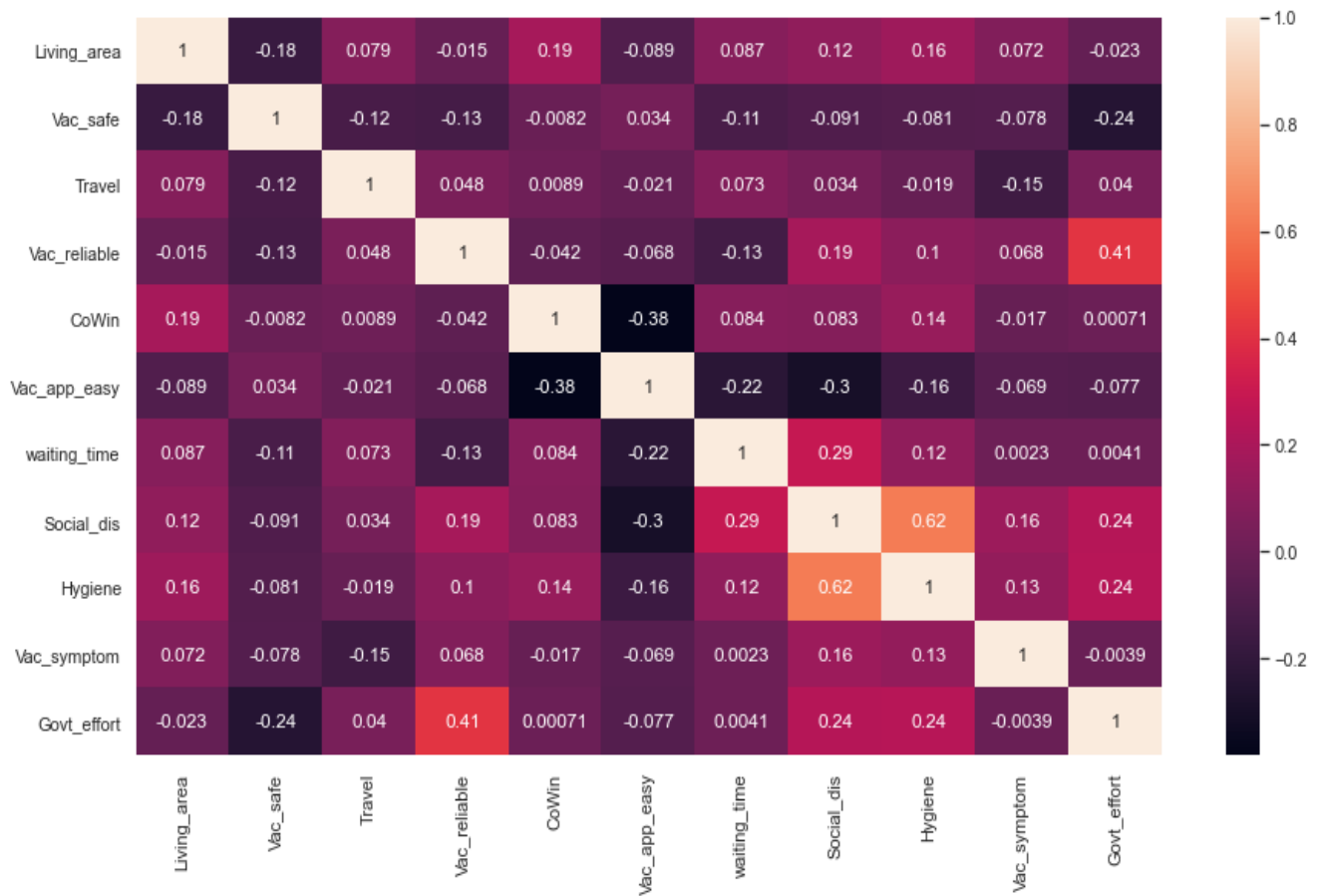
From the above pie diagram, the inferences we can make is,

- Only a few people aren't satisfied with the government's effort to meet the vaccination demand. A large chunk of people has rated the government effort above or equal to 3, thus showing more than average satisfaction regarding the government's effort to meet the vaccination demand.

From the pie Donut diagram, the inference we can draw,

- Only a small portion of the people of the age group of 18-25 and 26-35 have given a rating of 1, which resembles that they are not at all satisfied with the effort made by the government.
- All the people in age group of 45 and above seem to overall like the government's effort to meet the vaccination demand.
- Responses from the age group 36-45, seem to like the government's effort and give them a maximum rating of 5.

21.



The above correlation heatmap represents the spearman's rank correlation coefficients among the ranked columns. The columns which had string values, we have factorized those values and considered them as ranked data. There are few points we can observe from the heatmap,

- The correlation between vaccine reliability (private or Govt.) and Govt. effort is positive and moderately strong. That means people who rely more on Govt. provided free vaccines are more likely to believe that the Govt. has done a pretty good job Controlling the covid situation.
- People from more developed area are more sceptical about vaccine being safe.
- People who are satisfied with the Hygiene maintenance in the vaccination centre are observed to be satisfied with the social distance practice also because the coefficient is positive and mildly strong.



- People who are satisfied with the hygiene and social distancing in the vaccination centre are more prone to rate Govt. effort satisfactory.
- The coefficient between ease in getting an appointment and using CoWin portal to book the appointment tells us an interesting point. The people who used CoWin portal were those who faced most difficulty in getting the appointment and are more prone to give negative feedback about ease in getting an appointment.
- The people who booked their appointment easily are found to be more dissatisfied with the waiting time, hygiene and social distancing in queue in the vaccination centre.

# CONCLUSION

In the light of the collected data, we can conclude these things below,

- The elderly population is most likely to be sceptical about the impact of vaccines, whereas the young population in the age group 18-25 are most progressive about vaccination.
- Responders from rural areas are most progressive about vaccination, whereas people living in small town are seen to be most sceptical about it.
- Male responders are slightly more sceptical about the impact of vaccination than female responders.
- Responders from rural area are mostly inclined towards free vaccines given by the Govt., whereas people living in metropolitan cities are going for the paid vaccines given by some private entity.
- Vaccine availability is worst in rural areas and best in metropolitan cities.
- Most of the people who used CoWin portal to book a slot for vaccination faced a lot of difficulties.
- Young population is more prone to use CoWin portal for booking a slot and did not face any issue, whereas the people with age more than 36 faced a lot of issue handling it and most of them were refrained from even using it.
- People from rural and small town used the CoWin portal proportionally higher than people from metropolitan cities, which is pretty surprising.
- Covishield is the most preferred choice of vaccine brand across all the age groups and gender, whereas Moderna became the least preferred brand.
- People getting Covishield suffered from post vaccination symptoms mostly after the 1<sup>st</sup> dose, whereas people getting Covaxin are prone to get the symptoms after the 2<sup>nd</sup> dose. Also, most of people who received Covaxin did not suffer from any post vaccination symptoms at all!
- The young population of age group 18-25 have suffered the most after the 1<sup>st</sup> dose, whereas people in the age group 36-45 have suffered the most after both the doses. The elderly population is more probable to have no symptoms at all after the vaccination.
- Male responders suffered more from post vaccination symptoms than female in general.
- Feedback on waiting time in queue at the vaccination centre is best satisfactory for metropolitan cities and worst for small towns.
- Feedback on hygiene maintained at the vaccination centre is best satisfactory for rural areas and worst for metropolitan cities.
- Feedback on social distancing practiced at the vaccination centre is best satisfactory for small towns and worst for metropolitan cities.
- From the feedbacks on the vaccination centres depending on the types of the resident areas, we can say that vaccination centres in metropolitan cities have failed to maintain a hygienic and safe environment to conduct vaccination for public, whereas the people in rural areas are pretty satisfied with their vaccination centres.

# LIMITATIONS OF THIS PROJECT WORK

- Continuous data not included- Except from the few Likert scale, we didn't have any continuous data in our survey. Though this made the data quite easy to work with, having a continuous data would have added a new dimension to our EDA.
- Place of vaccination should have been included- Though it was asked that which vaccine would the respondent prefer, free vaccines by government institutions or paid vaccine by private nursing homes, it would have been a better choice to ask the respondent about the vaccine he has taken.
- Income- Should have taken income as a data point to make the analysis even better, because then we could have correlated the income groups with the inclination towards paid vaccines.
- Place of vaccination- Though the living area of the individual was taken into consideration, the actual place of vaccination wasn't taken into consideration. Adding such a data to our survey would help us in comparing the hygiene, social distancing and waiting time of the different living areas (rural, small town, metro).
- 0-17 age group- After data collection, we saw that there was only 1 data point in this age group which is not the ideal case. Also, the fact that the vaccines haven't been started for these age group of people, should have been added to the survey.

# ACKNOWLEDGEMENT

We would like to convey our sincere gratitude to our supervisor and mentor Prof. Sudeep Mallick for continuously sharing his expertise and helping us throughout the whole process whenever we needed any guidance. We would not be able to finish this project work if he did not help us with ideas and study materials constantly.

We would also like to extend our thanks to all of my other professors of the Department of Computer science, RKMVERI for their encouragement and kind co-operation.

Last but not the least, our special thanks go to our parents and classmates for staying by our side in this process and helped to keep our mind straight and focused.

# BIBLIOGRAPHY

Here are some of the references we used for our project-

- (Statistics and Computing) Peter Dalgaard (auth.) - Introductory Statistics with R- Springer-Verlag New York (2008)
- [R for data science by Hadley Wickham & Garreth Grolemund](#)
- [An Introduction to R- W. N. Venables, D. M. Smith and The R core Team](#)
- [R in a Nutshell- Book by Joseph Adler](#)
- [Exploratory Data Analysis with R by Roger D. Peng](#)
- [Python Data Science Handbook: Essential Tools for Working with Data](#)
- [Python for Data Analysis: Data Wrangling with Pandas, NumPy, and IPython](#)
- [Python Cookbook, 3rd Edition](#)

# APPENDIX

The questionnaire as a google form, used to collect the data is given in the link below,

GOOGLE FORM LINK - <https://forms.gle/HbsT79BxLhKUSy9U9>

The R-script containing all the codes used in this project, the dataset (in .csv format) is given below as a Github repository link,

GITHUB REPOSITORY LINK – [click here](#)