MSM Assignment-2 REPORT

Based on NFHS 2019-21 Data

Dhruv Choudhary	**2021B3A73142H**
Jyotirmoy Singh	**2021B3A72513H**
Sthitaprajna	**2021B3A71082H**
Aditya Kumar Jain	**2021B3A82573H**
Mugdha Gupta	**2021B3A72724H**
C Varun	**2021B3AA3031H**
Ananya Agrawal	**2021B3AA1935H**
Rachit Vaghani	**2021B3AA1457H**

Q1 : Write a brief introduction (300-400 words) of the datasets i.e. : NFHS

The fifth survey in the series, the National Family Health Survey 2019–21 (NFHS-5), offers data on India's population, health, and nutrition for each state and union territory (UT). Similar to NFHS-4, NFHS-5 offers district-level estimates for a number of significant metrics. The NFHS-5 sample has been created to offer estimates of several survey indicators at the national, state/union territory (UT), and district levels.Changes from NFHS-4 : waist and hip circumference measurements have been added to the list of clinical, anthropometric, and biochemical testing (CAB) procedures, and the age range for blood pressure and blood glucose measurements has been widened.

State/union territory (UT) along with federal figures are provided for indicators of sexual behaviour, husband`s background and wife`s employment, knowledge, attitudes, metrics related to HIV/AIDS, and domestic violence

The International Institute for Population Sciences, Mumbai, has been chosen by the Ministry of Health and Family Welfare of the Government of India as the nodal agency to conduct NFHS-5.

Along with the household`s socioeconomic characteristics, water, sanitation, and hygiene, health insurance coverage, disabilities, land ownership, the number of household deaths in the three years prior to the survey, and mosquito net ownership and use are major subtopics on which data was collected.

The **Woman`s Schedule** included information on a wide range of subjects, including the characteristics of women, marriage, fertility, contraception, **children`s vaccines** and healthcare, nutrition, reproductive health, sexual behaviour, HIV/AIDS, women`s empowerment, and domestic violence.

The Man`s Schedule included information on the man`s traits, marriage, number of children, use of contraception, preferences for fertility treatments, diet, sexual behaviour, health concerns, attitudes about gender roles, and HIV/AIDS. The Biomarker Schedule included measures of children`s height, weight, waist, and hip circumferences as well as

hemoglobin levels. It also included measurements of men and women's blood pressure as well as random blood glucose levels for those who were 15 years of age and older.

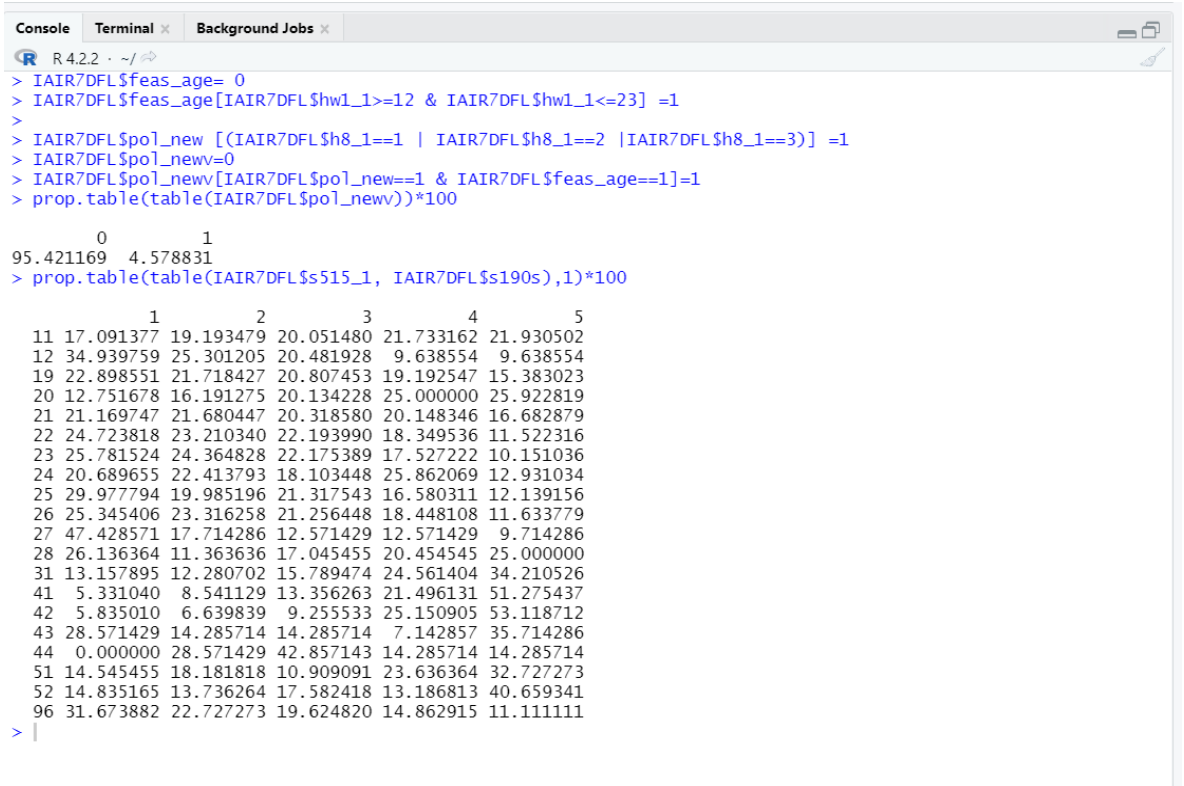
Information about significant trends and indicators for India is provided in a fact sheet. 17 Field Agencies carried out the NFHS-5 fieldwork for India in two stages, the first from 17 June 2019 to 30 January 2020 and the second from 2 January 2020 to 30 April 2021, collecting data from a total of 6,36,699 households, 7,24,115 women, and 1,01,839 men.

Q2 a :- Now carefully go through the variables list in modules selected. Based on your understanding of the same, you need to do simple tabulations, cross tabulations and calculate summary statistics of the variables your group deems important. Discuss the results.

INFERENCE:

s515 - Where receive most of his/her vaccinations?
s190s - Wealth index within state
Richest people constitute the biggest section of people to receive vaccines for their children from private hospitals.Most people from the the poor and poorest categories have got their children's vaccines from public PHC, sub-center, Anganwadi and camps. Poorest, poor, and middle class people's children got vaccines because of the effective implementation of the Anganwadi scheme. As a whole very few people opted for private Ayush for vaccination.

OUTPUT:



```
Console   Terminal ×   Background Jobs ×
R 4.2.2 · ~/
> IAIR7DFL$feas_age= 0
> IAIR7DFL$feas_age[IAIR7DFL$hw1_1>=12 & IAIR7DFL$hw1_1<=23] =1
>
> IAIR7DFL$pol_new [(IAIR7DFL$h8_1==1 | IAIR7DFL$h8_1==2 |IAIR7DFL$h8_1==3)] =1
> IAIR7DFL$pol_newv=0
> IAIR7DFL$pol_newv[IAIR7DFL$pol_new==1 & IAIR7DFL$feas_age==1]=1
> prop.table(table(IAIR7DFL$pol_newv))*100

       0        1
95.421169  4.578831
> prop.table(table(IAIR7DFL$s515_1, IAIR7DFL$s190s),1)*100

            1         2         3         4         5
  11 17.091377 19.193479 20.051480 21.733162 21.930502
  12 34.939759 25.301205 20.481928  9.638554  9.638554
  19 22.898551 21.718427 20.807453 19.192547 15.383023
  20 12.751678 16.191275 20.134228 25.000000 25.922819
  21 21.169747 21.680447 20.318580 20.148346 16.682879
  22 24.723818 23.210340 22.193990 18.349536 11.522316
  23 25.781524 24.364828 22.175389 17.527222 10.151036
  24 20.689655 22.413793 18.103448 25.862069 12.931034
  25 29.977794 19.985196 21.317543 16.580311 12.139156
  26 25.345406 23.316258 21.256448 18.448108 11.633779
  27 47.428571 17.714286 12.571429 12.571429  9.714286
  28 26.136364 11.363636 17.045455 20.454545 25.000000
  31 13.157895 12.280702 15.789474 24.561404 34.210526
  41  5.331040  8.541129 13.356263 21.496131 51.275437
  42  5.835010  6.639839  9.255533 25.150905 53.118712
  43 28.571429 14.285714 14.285714  7.142857 35.714286
  44  0.000000 28.571429 42.857143 14.285714 14.285714
  51 14.545455 18.181818 10.909091 23.636364 32.727273
  52 14.835165 13.736264 17.582418 13.186813 40.659341
  96 31.673882 22.727273 19.624820 14.862915 11.111111
>
```

Q2 b :-Generate two new variables (first variable with interval/ ratio scale of measurement and second with nominal/ ordinal scale of measurement) based on the existing variables. Discuss the steps involved. Apply the relevant descriptive statistics on the created variables.

INFERENCE:-

We created an interval type variable named “vac_sum”.

It is the sum of whether the person has taken Hepatitis B, BCG, Polio 1-3, Pentavalent 1-3, Measles, Vitamin A1, Vitamin A2, Measles vaccines. These particular 12 vaccines as directed by the National Immunization Schedule must be taken by the child. If the person has a perfect score of 12, it implies that the person has taken all 12. Ideally a score of 10 or greater is preferred.

We have created several nominal type variables for reference, the two that we used the most were:

“feas_age” and “q57_new”

As most of the vaccination questions are pertaining to children between the age of 12-23 months, we have created the feas_age variable for easier variable creation and access.

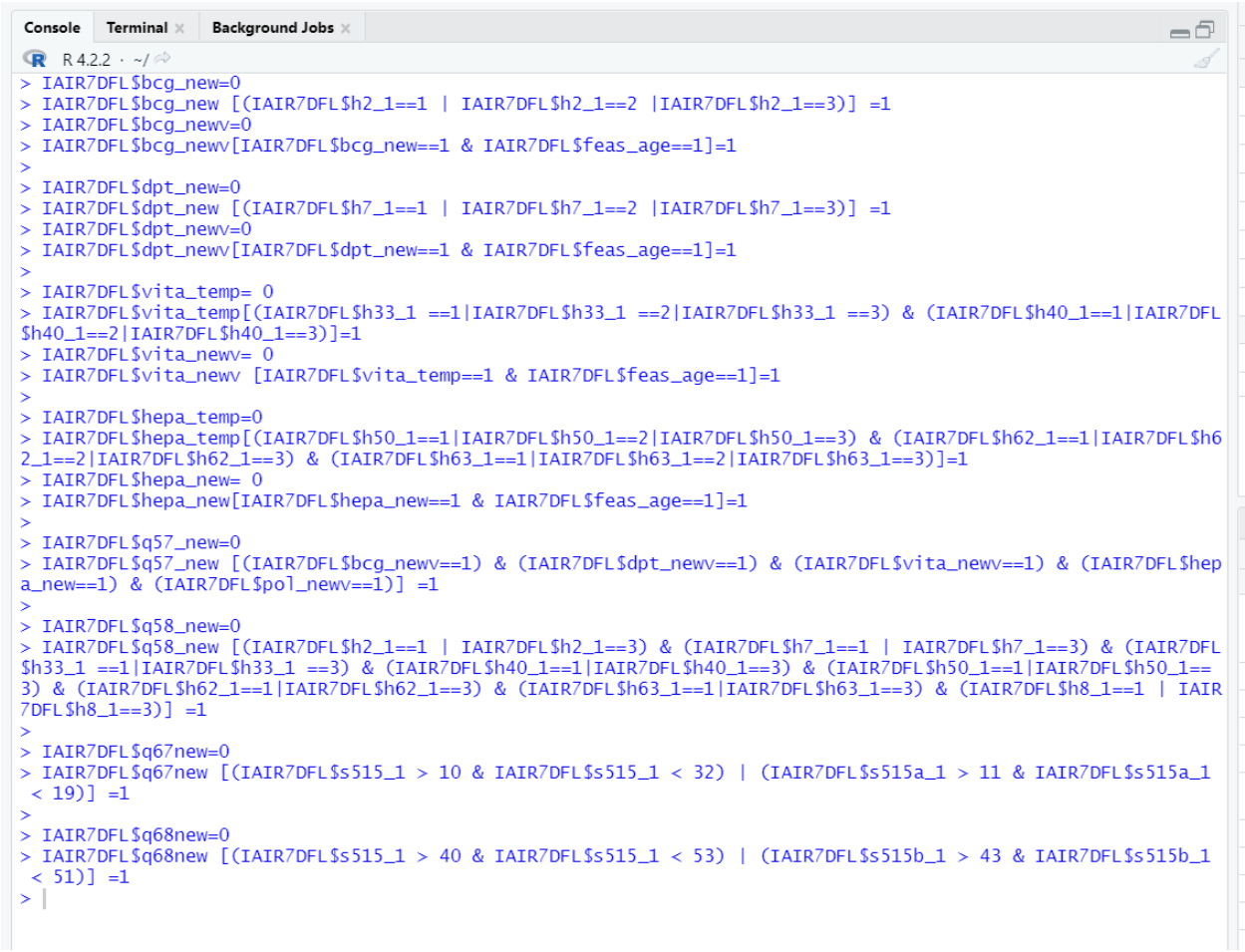
The variable q57_new answers the question: Whether the children between the age of 12-23 months are fully vaccinated or not. The answer is 0 by default, and 1 if yes. In order to create this variable, we have created several variables to check if they are fully BCG, DPT, Vitamin A, hepatitis B and Polio vaccinated based on their number of doses required. Following is the script for the same:

We created one to check whether third dose has been given for completely vaccinated and all

After applying relevant descriptive statistics on some created variables we found :-

- 94.67 % of the children age between 12-23 months were vaccinated for tuberculosis.
- 84.89% of the children took most of their vaccines in a public health facility.
- 94.62 % of the children age between 12-23 months were fully vaccinated for polio.

OUTPUT:-



```
Console  Terminal  Background Jobs
R 4.2.2 · ~/
> IAIR7DFL$bcg_new=0
> IAIR7DFL$bcg_new [(IAIR7DFL$h2_1==1 | IAIR7DFL$h2_1==2 |IAIR7DFL$h2_1==3)] =1
> IAIR7DFL$bcg_newv=0
> IAIR7DFL$bcg_newv[IAIR7DFL$bcg_new==1 & IAIR7DFL$feas_age==1]=1
>
> IAIR7DFL$dpt_new=0
> IAIR7DFL$dpt_new [(IAIR7DFL$h7_1==1 | IAIR7DFL$h7_1==2 |IAIR7DFL$h7_1==3)] =1
> IAIR7DFL$dpt_newv=0
> IAIR7DFL$dpt_newv[IAIR7DFL$dpt_new==1 & IAIR7DFL$feas_age==1]=1
>
> IAIR7DFL$vita_temp= 0
> IAIR7DFL$vita_temp[(IAIR7DFL$h33_1 ==1|IAIR7DFL$h33_1 ==2|IAIR7DFL$h33_1 ==3) & (IAIR7DFL$h40_1==1|IAIR7DFL
$h40_1==2|IAIR7DFL$h40_1==3)]=1
> IAIR7DFL$vita_newv= 0
> IAIR7DFL$vita_newv [IAIR7DFL$vita_temp==1 & IAIR7DFL$feas_age==1]=1
>
> IAIR7DFL$hepa_temp=0
> IAIR7DFL$hepa_temp[(IAIR7DFL$h50_1==1|IAIR7DFL$h50_1==2|IAIR7DFL$h50_1==3) & (IAIR7DFL$h62_1==1|IAIR7DFL$h6
2_1==2|IAIR7DFL$h62_1==3) & (IAIR7DFL$h63_1==1|IAIR7DFL$h63_1==2|IAIR7DFL$h63_1==3)]=1
> IAIR7DFL$hepa_new= 0
> IAIR7DFL$hepa_new[IAIR7DFL$hepa_new==1 & IAIR7DFL$feas_age==1]=1
>
> IAIR7DFL$q57_new=0
> IAIR7DFL$q57_new [(IAIR7DFL$bcg_newv==1) & (IAIR7DFL$dpt_newv==1) & (IAIR7DFL$vita_newv==1) & (IAIR7DFL$hep
a_new==1) & (IAIR7DFL$pol_newv==1)] =1
>
> IAIR7DFL$q58_new=0
> IAIR7DFL$q58_new [(IAIR7DFL$h2_1==1 | IAIR7DFL$h2_1==3) & (IAIR7DFL$h7_1==1 | IAIR7DFL$h7_1==3) & (IAIR7DFL
$h33_1 ==1|IAIR7DFL$h33_1 ==3) & (IAIR7DFL$h40_1==1|IAIR7DFL$h40_1==3) & (IAIR7DFL$h50_1==1|IAIR7DFL$h50_1==
3) & (IAIR7DFL$h62_1==1|IAIR7DFL$h62_1==3) & (IAIR7DFL$h63_1==1|IAIR7DFL$h63_1==3) & (IAIR7DFL$h8_1==1 | IAIR
7DFL$h8_1==3)] =1
>
> IAIR7DFL$q67new=0
> IAIR7DFL$q67new [(IAIR7DFL$s515_1 > 10 & IAIR7DFL$s515_1 < 32) | (IAIR7DFL$s515a_1 > 11 & IAIR7DFL$s515a_1
 < 19)] =1
>
> IAIR7DFL$q68new=0
> IAIR7DFL$q68new [(IAIR7DFL$s515_1 > 40 & IAIR7DFL$s515_1 < 53) | (IAIR7DFL$s515b_1 > 43 & IAIR7DFL$s515b_1
 < 51)] =1
>
```

- For BCG Vaccine (Var: bcg_newv) :-

```
> prop.table(table(IAIR7DFL$bcg_newv))*100

        0         1
94.671979  5.328021
```

- Vaccination in a public health facility (Var: q67_new) :-

```
> prop.table(table(IAIR7DFL$q67new))*100

       0        1
84.89135 15.10865
```

- For Polio Vaccine (Var: polio_newv) :

```
> prop.table(table(IAIR7DFL$pol_newv))*100

        0         1
94.625439  5.374561
```

Q2 c:- Use the newly created variable to do cross-tabulations across the states of India. Discuss your result.

INFERENCE:

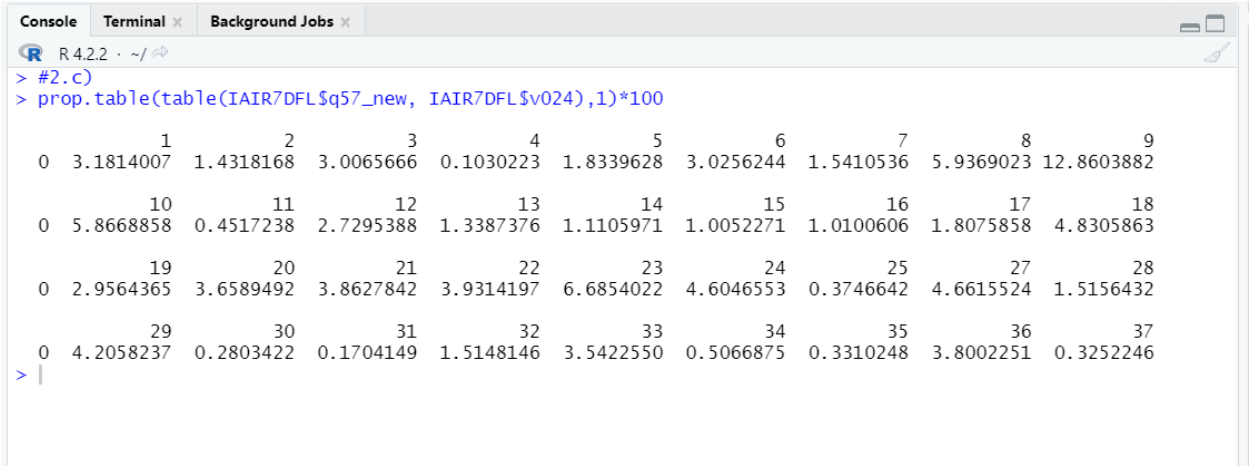
On cross tabulating the newly created variable q57_new with state, we get an insight on from which state the percentage of children who are fully immunized are from.

Looking at the data we can see that the most number of children who were fully immunized were from Uttar Pradesh at 12.86% , it being a state with more population and the lowest were from Lakshadweep at 0.17 %, it being a small UT.

States like Jammu & Kashmir, Punjab, Haryana, Rajasthan, Bihar, Assam, Jharkhand, Odisha, Chattisgarh, Gujarat, Maharashtra, Karnataka and Telangana contributed to 3-6% of the total children fully immunized.

States/UT like Chandigarh, Sikkim, Dadra & Nagar Haveli And Daman & Diu, Puducherry, Andaman & Nicobar Islands, Ladakh, Lakswadeep and Goa contributed 1% of the children fully immunized of the country.

Output:-

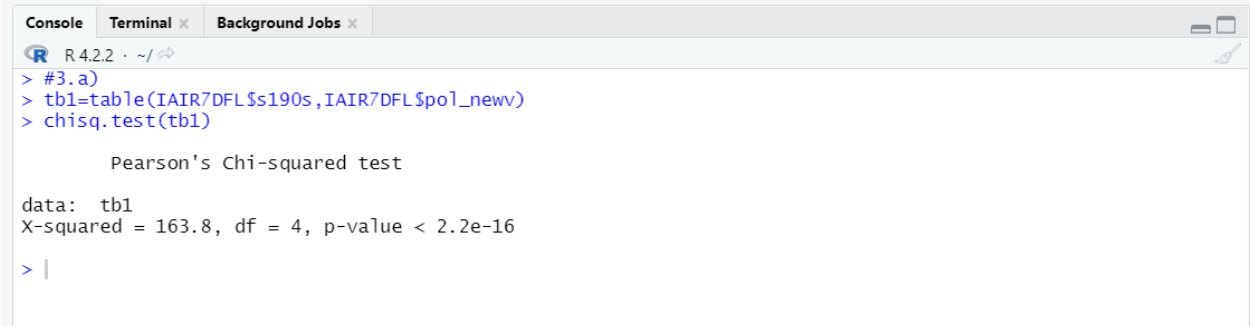


Q3 a:- Try to find association between relevant categorical variables and comment on the result.

INFERENCE:

Following the chi square test for the variables wealth index and the number of individuals who have received all doses of the polio vaccine, we find that the p value is very low and is less than alpha, which is set to 95% by default in R, so we can reject the null hypothesis. This indicates that the two variables are dependent on and related to one another.

OUTPUT:

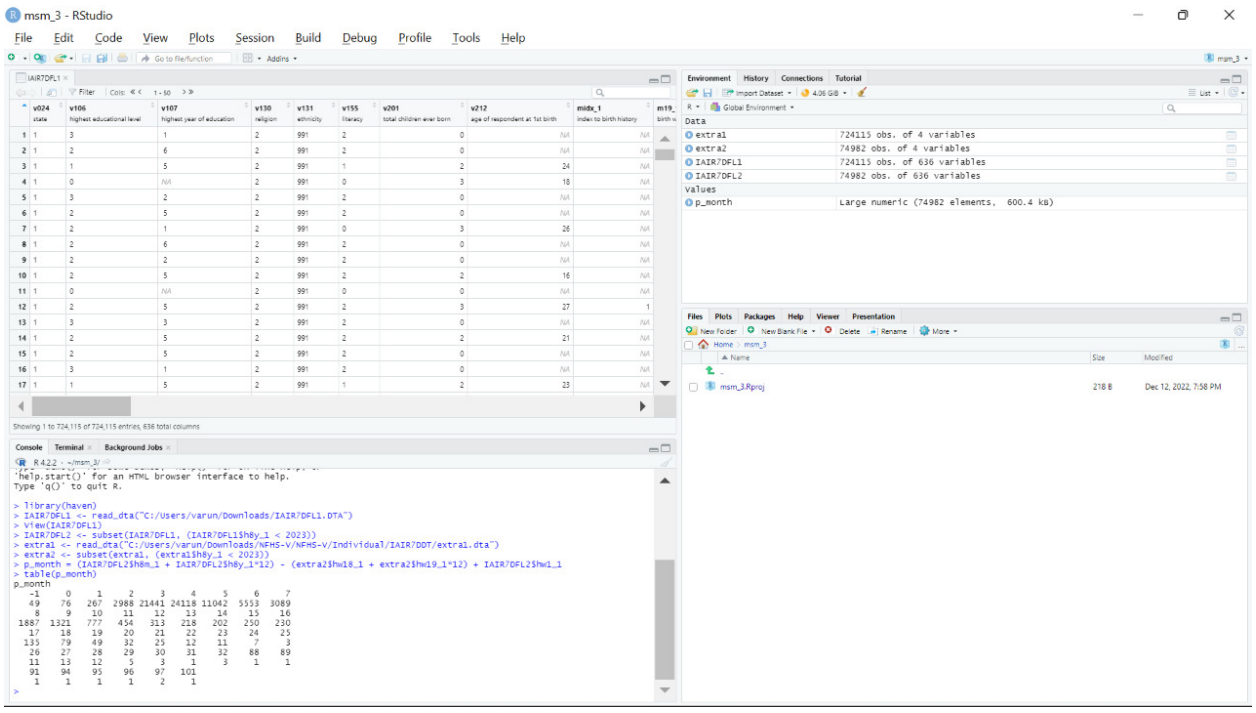


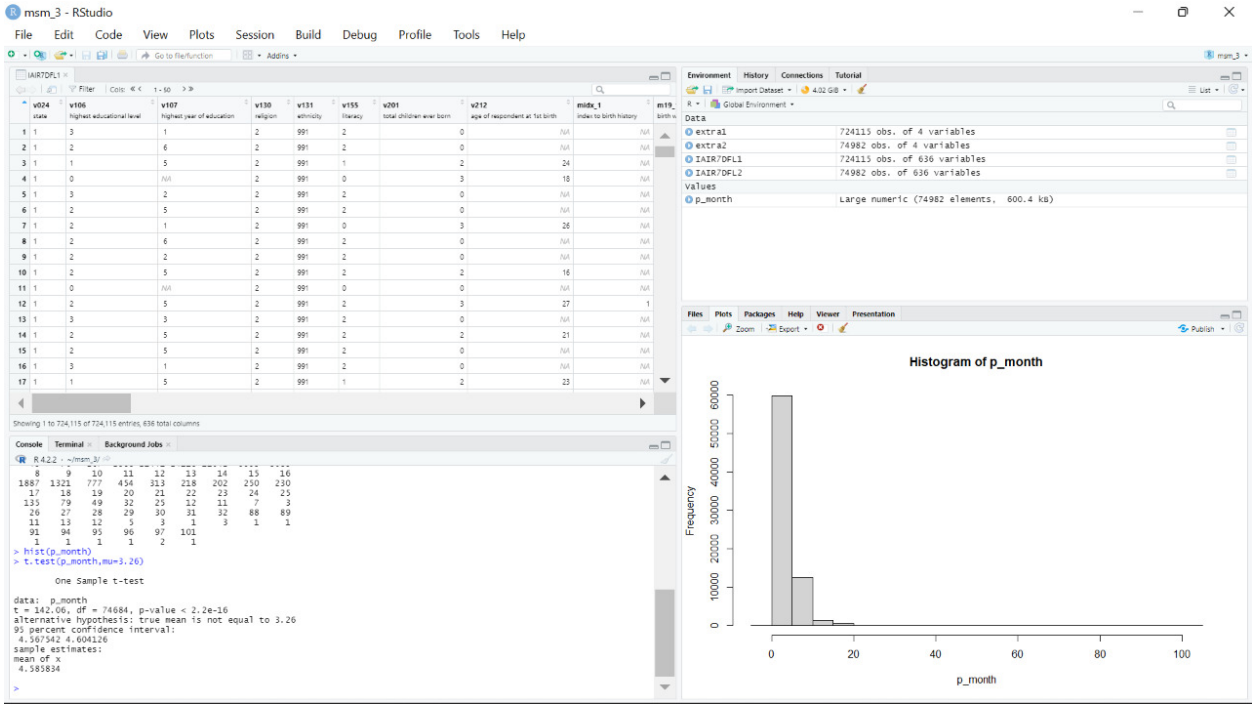
Q3 b:- Use any of the “t-tests” discussed to test any relevant hypothesis. Discuss the findings.

INFERENCE:

The variable used for t test is the month at which the child was given the third dose of vaccine. This, according to MoFHS and experts should be 14 weeks which translates to approximately 3.26 months. The value was calculated by adding the date the child was given a vaccine(in months) with the age of the child(in months) and subtracting it from the the date the interview was taken(in months). Finally, the variable that was created, was run a t test with mu = 3.26. The 95% confidence interval came out to be [4.5675 4.6041]. This means that the mu value significantly deviates from 3.26 and hence null hypothesis has to be rejected.

OUTPUT:




msm_3 - RStudio
File Edit Code View Plots Session Build Debug Profile Tools Help
Environment History Connections Tutorial
Data
extra1 724115 obs. of 4 variables
extra2 74982 obs. of 4 variables
IAIR7DFL1 724115 obs. of 636 variables
IAIR7DFL2 74982 obs. of 636 variables
Values
p_month Large numeric (74982 elements, 600.4 kB)
Files Plots Packages Help Viewer Presentation
Histogram of p_month
Frequency
p_month
> hist(p_month)
> t.test(p_month,mu=3.26)
One Sample t-test
data: p_month
t = 142.06, df = 74984, p-value < 2.2e-16
alternative hypothesis: true mean is not equal to 3.26
95 percent confidence interval:
4.567542 4.604126
sample estimates:
mean of x
4.585834

Q3 c:- Use the newly created variable to do cross-tabulations across the states of India. Discuss your result.

INFERENCE:

The Regression Model obtained is:
finreg= lm(IAIR7DFL$q67new~IAIR7DFL$v107+ extra_1$v102)

Here the Dependent variable(Y) is the newly created variable (in 2b), q67new which is a nominal variable denoting whether the person is fully vaccinated or not. The independent variable (x1) is the number of years of education . The second independent variable (x2) is the wealth index factor score (5 decimal places).

We can represent a general regression equation as:
$\mathbf{Y = b_0 + b_1x_1 + b_2x_2}$
Here b_0 = 0.0517 while b_1= 0.00405 and b_2= 0.0512.

This denotes that the graph has an intercept of 0.0517. This implies if the wealth index is 0 and the number of years of education is 0, the probability of the person being fully vaccinated is 0.0517.

b_1= 0.00405 denotes that with a unit increase in years of education, the probability of the person being fully vaccinated increases by 0.00405.
b_2= 0.0512 denotes that with a unit increase in wealth index, the probability of the person being fully vaccinated increases by 0.0512.

OUTPUT:

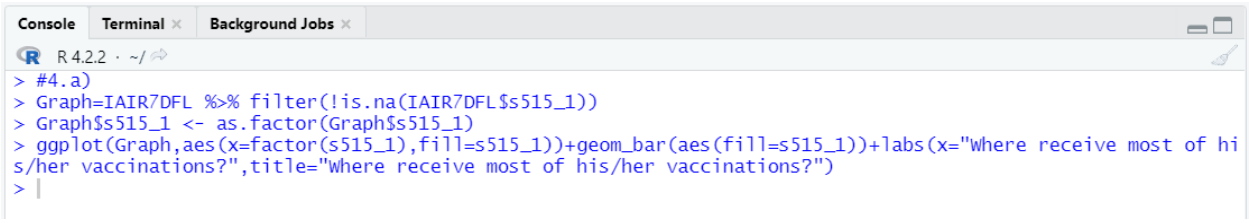
finreg	list [13] (S3: lm)	List of length 13
coefficients	double [3]	0.05178 0.00405 0.05216
residuals	double [556811]	-0.160 -0.180 -0.176 -0.164 -0.176 -0.160 ...
effects	double [556811]	-118.131 -4.432 -17.496 -0.164 -0.176 -0.160 ...
rank	integer [1]	3
fitted.values	double [556811]	0.160 0.180 0.176 0.164 0.176 0.160 ...
assign	integer [3]	0 1 2
qr	list [5] (S3: qr)	List of length 5
df.residual	integer [1]	556808
na.action	integer [167304] (S3: omit)	4 11 18 23 26 28 ...
xlevels	list [0]	List of length 0
call	language	lm(formula = IAIR7DFL$q67new ~ IAIR7DFL$v107 + extra_1$\
terms	formula	IAIR7DFL$q67new ~ IAIR7DFL$v107 + extra_1$v102

Q4 a:- Use ggplot2 package to make a plot of any set of relevant variables. Discuss the results

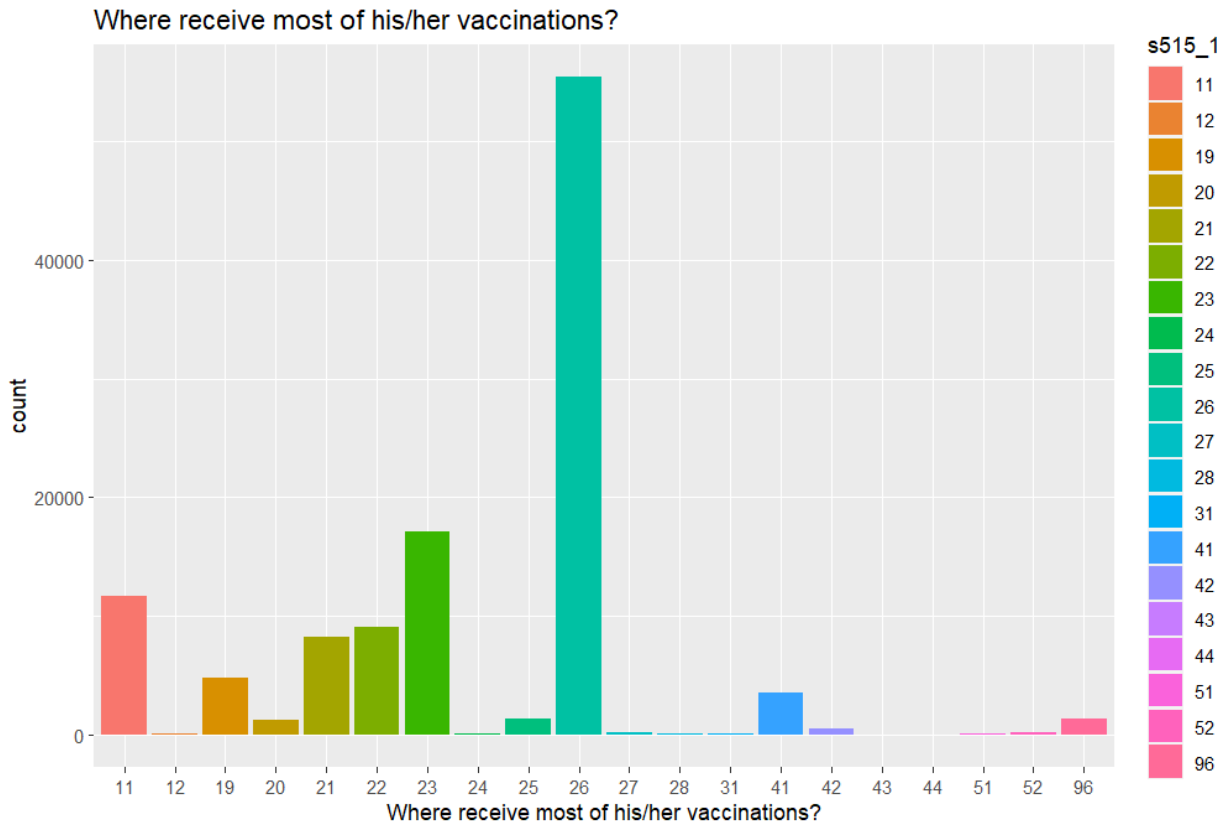
INFERENCE:

We have made a bar graph showing the places where the respondent's child received most of their vaccinations .From the graph we can see that most of the people took their vaccines from public areas compared to private. We can also see that most of the people opted for Public Anganwadi and ICDS centre for their vaccination doses , this shows that the Anganwadi program set up by the government is proving to be more effective as it is now properly catering to the physical development of children.The private sector plays a comparatively minor role in the process.The immunization process, can aim to accommodate more participation from NGO's.

SCRIPT:



OUTPUT:



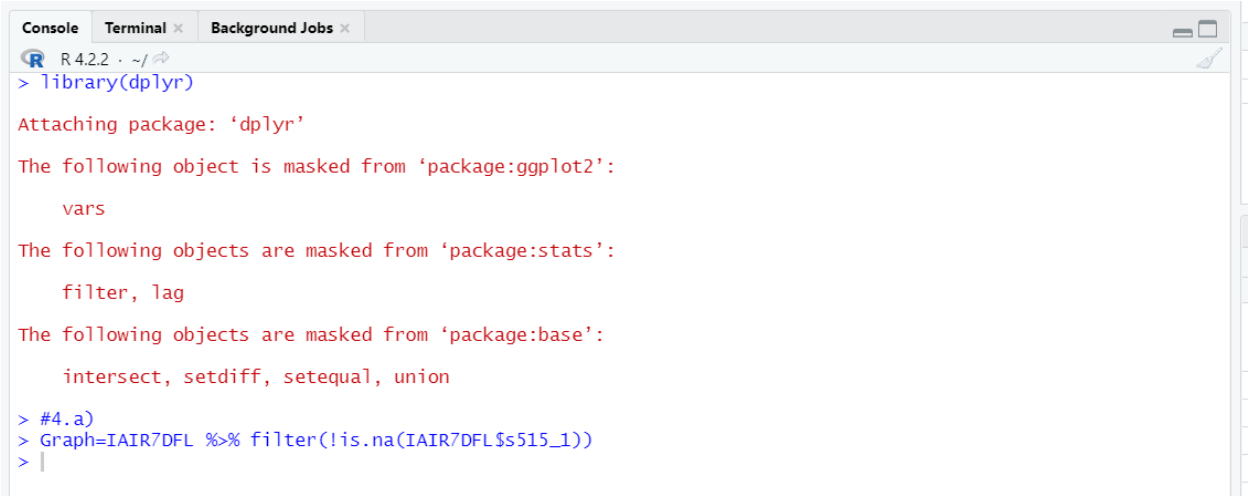
Response	Label	Response	Label
11	Public: Govt. / Municipal Hospital	27	Public: PULSE Polio
12	Public AYUSH	28	Other Public Sector Health Facility
19	Public: Govt. Dispensary	31	NGO or Trust Hospital / Clinic
20	Public: UHC / UHP / UFWC	41	Private: Hospital

21	Public: CHC / Rural Hospital / Block PHC	42	Private: Doctor / Clinic
22	Public: PHC / Additional PHC	43	Private: Paramedic
23	Public: Sub-Centre / ANM	44	Private AYUSH
24	Public: Govt. Mobile Clinic	51	Private: Pharmacy / Drugstore
25	Public: camp	52	Other Private Health Faility
26	Public: Anganwadi / ICDS Centre	96	Other

Q4 b. Use any other relevant R package on your data and explain the results.

We used the package “dplyr” to remove all the NA’s from a specific variable in our dataset so we could run the other commands properly without any error.This package enabled us to use the function “%>%” to remove the NA’s from the variable s515_1 which is the variable for “Where receive most of his/her vaccinations?”.

We ran the following command :



This enabled us to completely remove the rows of observations from the dataset where the value of the variable s515_1 was NA.This helped us to run the “ggplot” command to make a bar graph which was earlier not possible properly due to the presence of NA’s.