

E-CIGARETTES AND THE INDIAN YOUTH: A STATISTICAL STUDY

SHRUTIKA GUPTA

ROLL NO. 0472

Under the supervision of

DR. SURUPA CHARABORTY (ROY)



Department of Statistics

St. Xavier's College (Autonomous), Kolkata

WHAT ARE E-CIGARETTES?

- Battery-operated rechargeable sleek smoking devices
- contain nicotine and carcinogens
- commonly known as 'vaping'
- Government ban under PECA in 2019



POPULATION

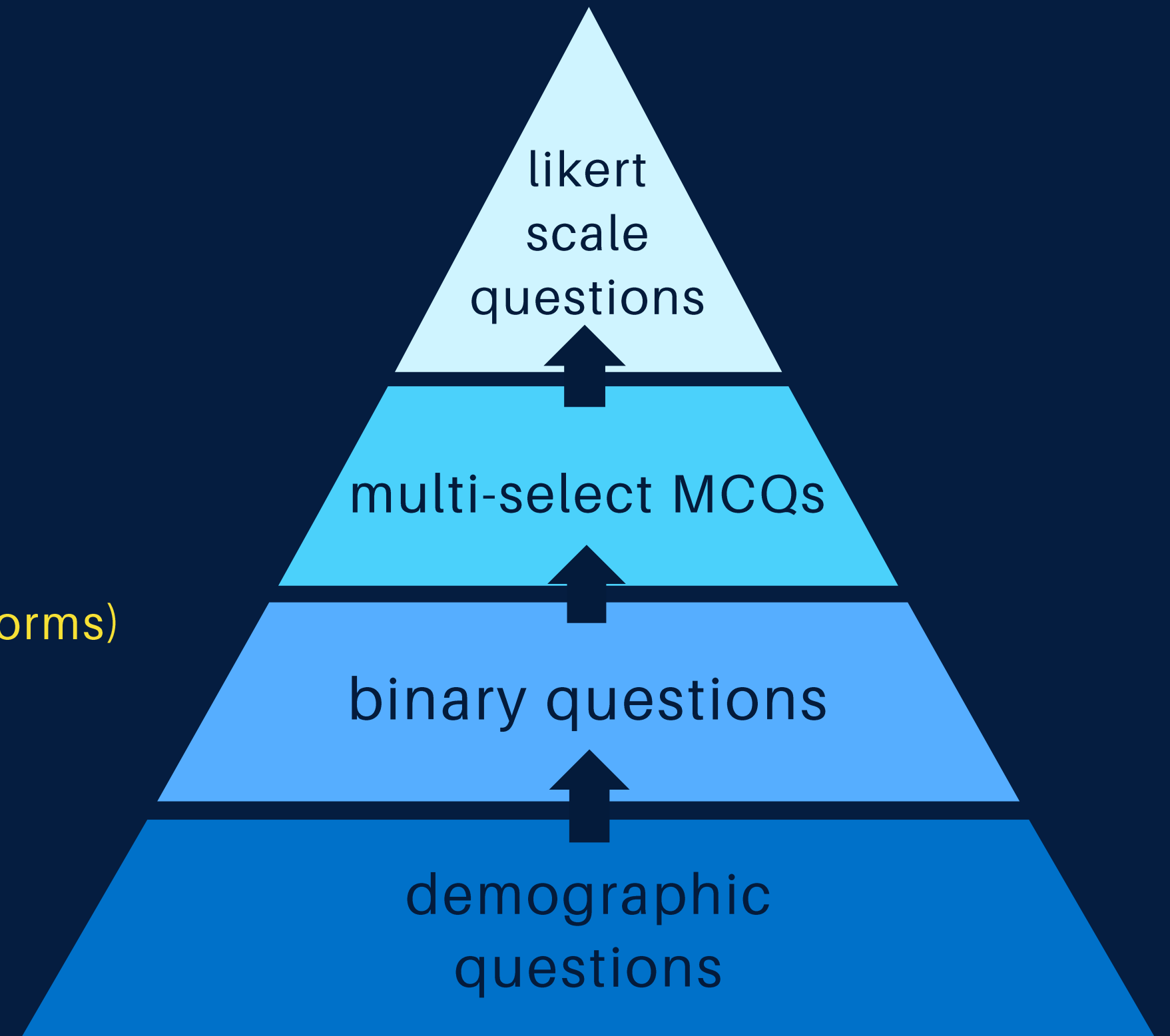
- College students
- Ages: 18–22
- Annual income between INR 20–50L
- From Kolkata, Delhi, and Mumbai



METHODOLOGY

- primary data
- cross-sectional data
- through an anonymous online survey (Google Forms)

Types of questions



QUESTIONNAIRE

QUESTIONNAIRE

Ethical Considerations: It is to be noted that this survey has been conducted following all ethical norms. Informed consent has been obtained from each voluntary participant prior to beginning the survey. Their identities will be kept anonymous and their responses confidential to ensure an entirely unbiased study.

1. How old are you ?

- ☐ 18
- ☐ 19
- ☐ 20
- ☐ 21
- ☐ 22

2. What is the gender that you identify as ?

- ☐ Male
- ☐ Female
- ☐ Other

3. What is your annual family income (approximately, in INR) ?

- ☐ Less than 20L
- ☐ 20-30L
- ☐ 30-40L
- ☐ 40-50L
- ☐ More than 50L

4. Have you ever smoked a cigarette ?

- ☐ Yes
- ☐ No

5. How many friends do you have who smoke e-cigarettes ?

- ☐ 0
- ☐ 1
- ☐ 2
- ☐ 3 or more

6. Have you ever smoked an e-cigarette ? (Note: even once or twice counts as a ‘yes’)

- ☐ Yes
- ☐ No

Section (I): For those who answered ‘NO’ to Question 6

7. Would you ever try an e-cigarette out of curiosity ?

- ☐ Definitely
- ☐ Maybe
- ☐ Never

8. Would you ever try an e-cigarette if your friend(s) offered you one ?

- ☐ Definitely
- ☐ Maybe
- ☐ Never

Section (II): For those who answered ‘YES’ to Question 6

9. How did you first get introduced to e-cigarettes? (choose all that apply)

- ☐ First saw/heard friends using them
- ☐ Saw people using them on social media platforms
- ☐ Wanted to try them out of curiosity
- ☐ Friends/peers offered one to try
- ☐ Wanted an alternative to smoking regular cigarettes

10. How did you gain access to e-cigarettes ? (choose all that apply)

- ☐ Borrowed from friends/peers
- ☐ Purchased from vape shops (online/offline)

11. Where have you seen e-cigarettes being shown/advertised the most ? (choose all that apply)

- ☐ Social media
- ☐ Movies/TV shows
- ☐ Fliers and/or other forms of print media

12. When do you smoke e-cigarettes ? (choose all that apply)

- ☐ When with friends (at parties/restaurants etc)
- ☐ When alone (at home or elsewhere)
- ☐ When feeling stressed/nervous

13. How long have you been smoking e-cigarettes ?

- ☐ Less than 1 year
- ☐ 1-2 years
- ☐ 2-3 years
- ☐ More than 3 years

14. What are your reasons for continuing to vape ? (choose all that apply)

- ☐ Sweet fruity taste/fresh flavours
- ☐ Friends do it
- ☐ Ease of availability (my friends and/or I have our own vapes)
- ☐ Family does not know yet
- ☐ I use it to try to quit smoking

15. Indicate how far you agree/disagree with the statements given below.

Statements	Strongly Agree	Agree	Neutral	Disagree	Strongly Disagree
E-cigarettes are addictive					
E-cigarettes contain nicotine					
E-cigarettes are better than tobacco cigarettes					
E-cigarette packages should come with clear health warnings					
Misleading advertisements that claim e-cigarettes to be ‘safe’ or advertisements with celebrities should be prohibited as they have the ability to sway public opinion					

Section (III): For those who smoke tobacco cigarettes.

16. Do you prefer e-cigarettes to regular cigarettes ?

- ☐ Yes
- ☐ No
- ☐ Not a smoker

17. Do you use e-cigarettes to give up smoking regular cigarettes ?

- ☐ Yes
- ☐ No
- ☐ Not a smoker

Section (IV): Question regarding awareness regarding ban on e-cigarettes

18. Did you know that under the Prohibition of Electronic Cigarettes Act (PECA), 2019, the production, manufacture, import, export, transport, sale, distribution, storage and advertisement of e-cigarettes is banned in India ?

- ☐ Yes
- ☐ No

DESCRIPTION OF THE DATASET

SAMPLE SIZE 252

RESPONSE VARIABLE (BINARY): y denoting the vaping status of a person, where $y = 1$ for yes
0 for no

EXPLANATORY VARIABLES : $x_1, x_2, x_3, x_4, x_5,$

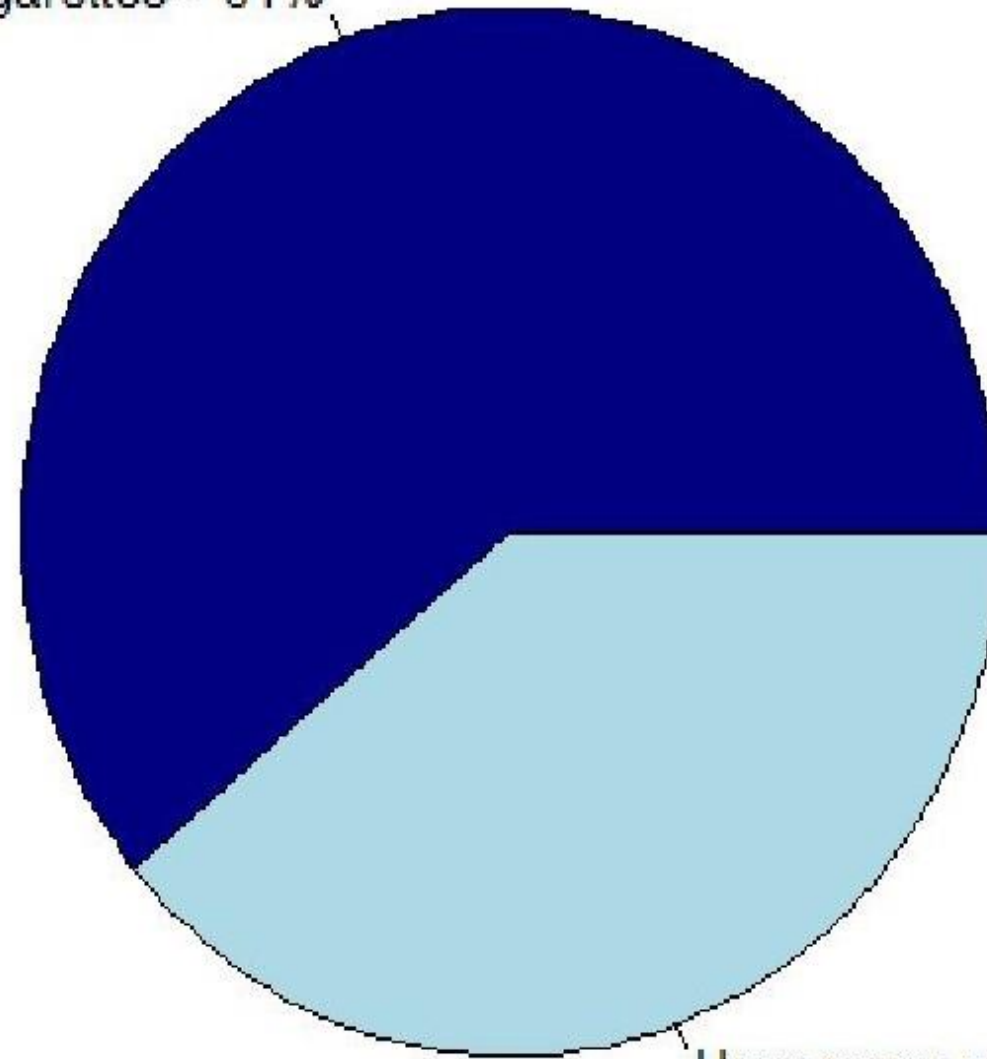
where,

- x_1 : the age of a person (continuous) i.e. 18 - 22 years
- x_2 : the gender of a person (binary),
 $x_2=1$ for male
0 for female
- x_3 : the annual income of a person (binary),
 $x_3=1$ for high income class ($>40L$)
0 for low income class ($<40L$)
- x_4 : the smoking status of a person (binary),
 $x_4=1$ for yes
0 for no
- x_5 : the number of friends of the respondent who vape,
 $x_5=1$ for ≥ 1 such friends
0 for no such friends

DATA VISUALIZATION

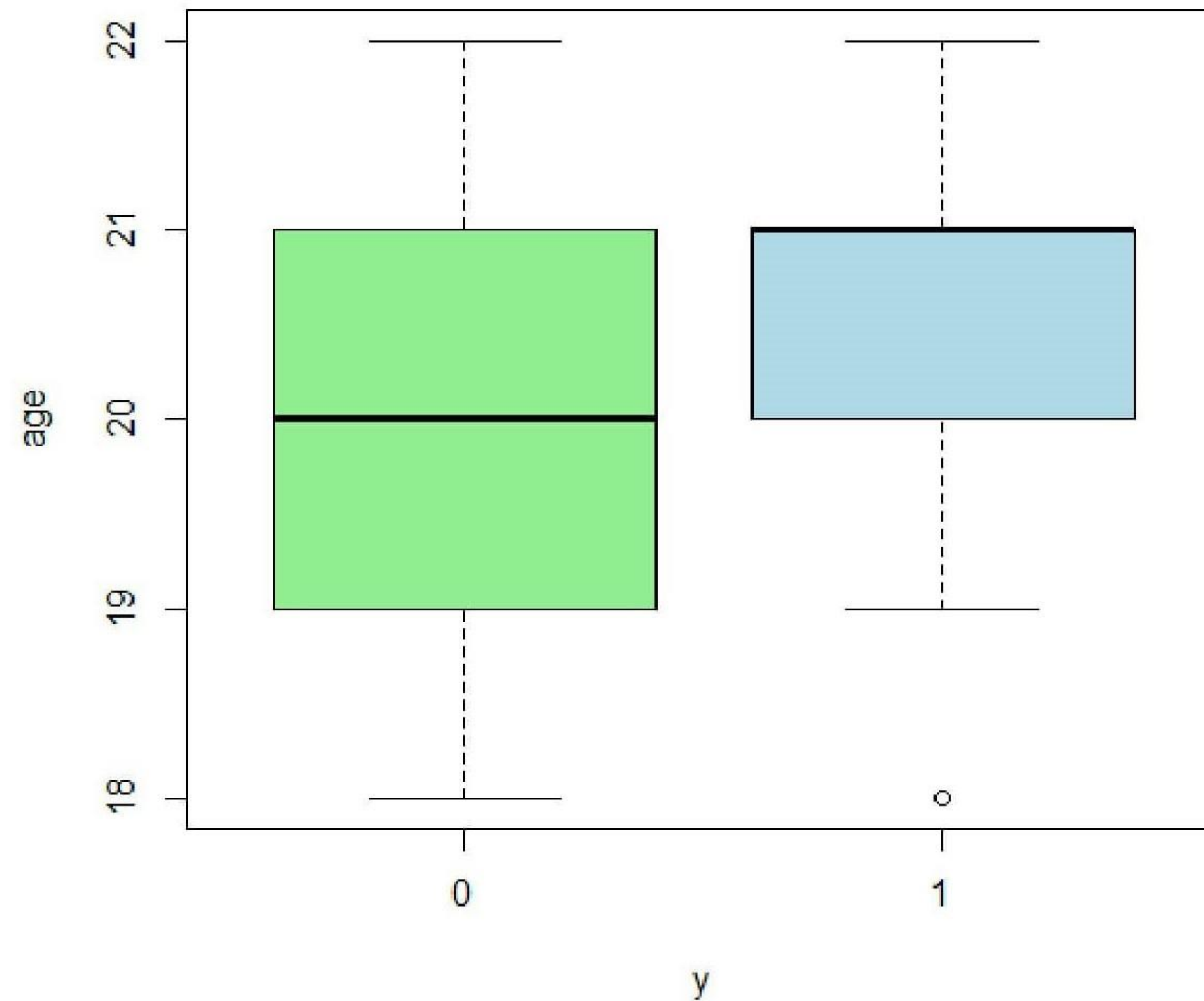
Pie Chart of Proportion of Respondents who Consume E-Cigarettes

Have smoked e-cigarettes - 61%

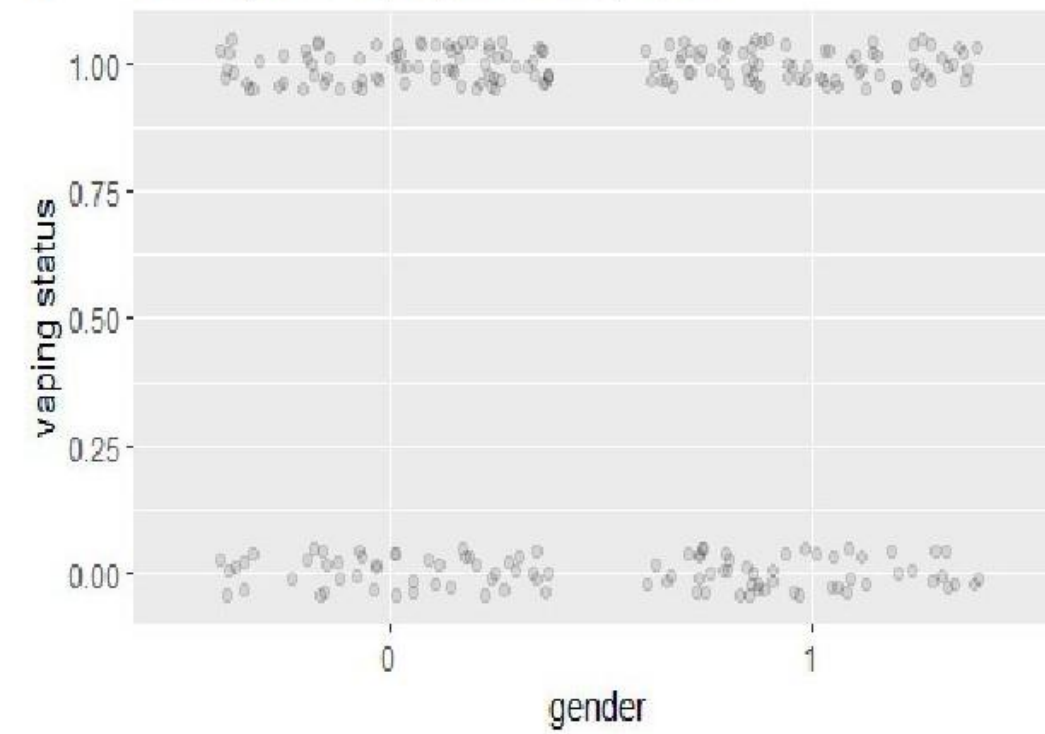


Have never smoked e-cigarettes - 39%

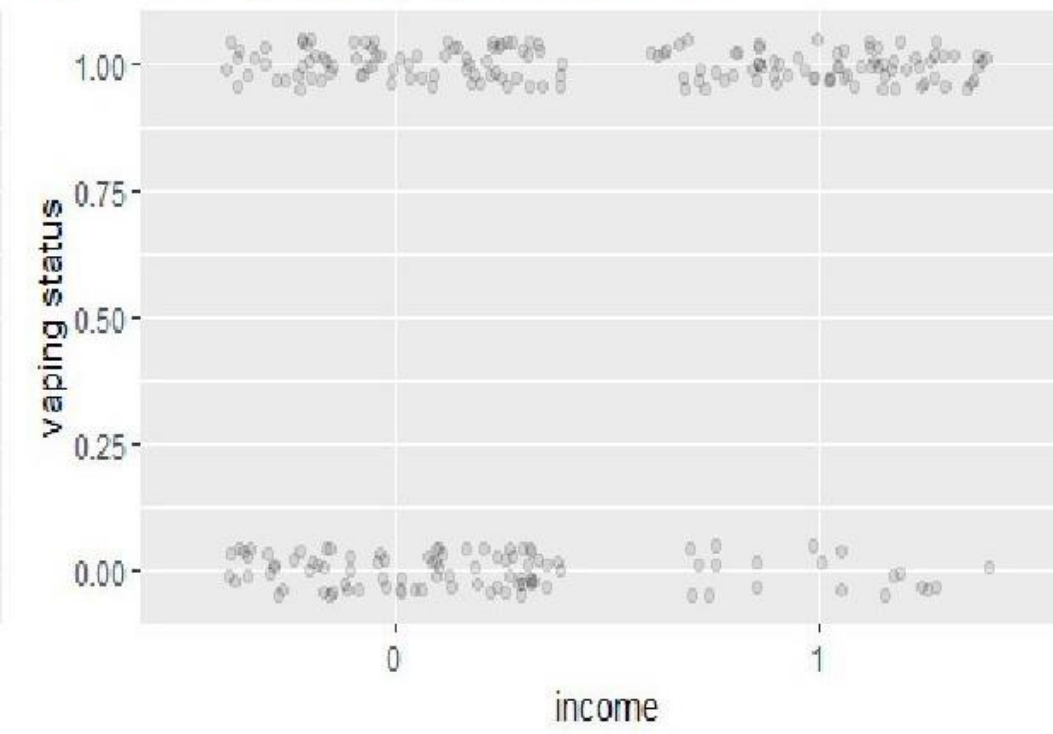
Boxplots of Ages for Different Responses ($y=0$ and $y=1$)



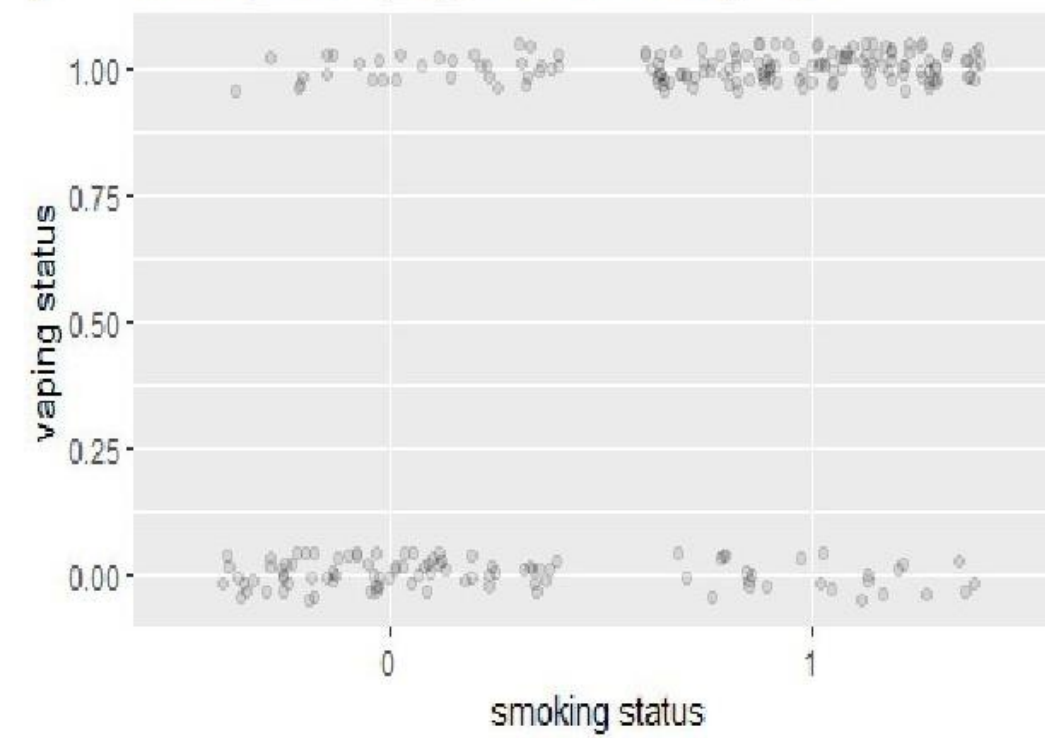
A Scatterplot of vaping status vs gender



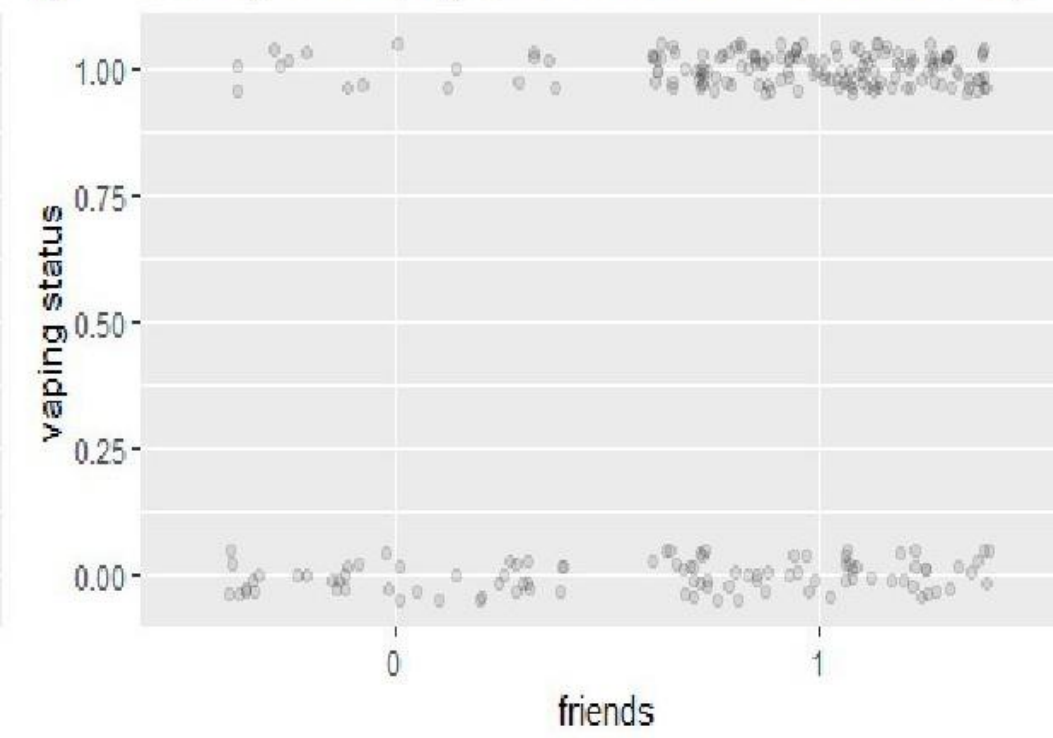
B Scatterplot of vaping status vs income



C Scatterplot of vaping status vs smoking status



D Scatterplot of vaping status vs number of friends who vape



LOGISTIC REGRESSION

A multivariable logistic regression model was fit to the data to identify the factors most closely associated with e-cigarette use.

Let π denote the probability of success i.e. $y=1$, such that $0 < \pi < 1$.

Hence, we can say,

$$y \sim \text{Bernoulli}(\pi)$$

Which means $E(y) = \mu = \pi$.

In particular, $P(y=1 \mid x_1, x_2, x_3, x_4, x_5) = \pi$ and $P(y=0 \mid x_1, x_2, x_3, x_4, x_5) = 1-\pi$

Model: The model under consideration is then given as:

$$E(y \mid x_1, x_2, x_3, x_4, x_5) = \eta = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \beta_3 x_3 + \beta_4 x_4 + \beta_5 x_5$$

Where $(x_1, x_2, x_3, x_4, x_5)$ is a set of explanatory variables and $(\beta_0, \beta_1, \beta_2, \beta_3, \beta_4, \beta_5)$ are unknown regression parameters that are to be estimated, and η is a linear predictor.

Here, we may consider the **Logit Link** function, that is also the canonical link.

$$\text{i.e.,} \quad \pi = \frac{e^{\beta_0 + \beta_1 x_1 + \beta_2 x_2 + \beta_3 x_3 + \beta_4 x_4 + \beta_5 x_5}}{1 + e^{\beta_0 + \beta_1 x_1 + \beta_2 x_2 + \beta_3 x_3 + \beta_4 x_4 + \beta_5 x_5}} \quad \text{--- (ii)}$$

Equation (ii) is considered the **multiple logistic regression equation**, using the logit link function.

TABLE OF CALCULATIONS

To test:

$H_0: \beta_0 = \beta_1 = \beta_2 = \beta_3 = \beta_4 = \beta_5 = 0$ against $H_1: \text{not } H_0$ (ie. at least one of the equalities do not hold true)

Test Statistic: Under H_0 , $Z_j = \frac{\widehat{\beta_j}}{SE(\widehat{\beta_j})} \sim t_{n-2} = t_{250} \quad \forall j=0(1)5$

where $\widehat{\beta_j}$ is the estimate of the coefficient of the jth predictor, and $SE(\widehat{\beta_j})$ is its standard error, $\forall j=0(1)5$.

Test Rule: We reject H_0 at $\alpha=0.05$ if the p-value $\Pr(> |Z_j|)$ is less than α i.e. if

$$\Pr(> |Z_j|) < \alpha = 0.05$$

Table of Calculations:

$\widehat{\beta_j}$	$SE(\widehat{\beta_j})$	Z_j (obs.)	$\Pr(> Z_j)$	Decision
$\widehat{\beta_0} = -1.48870$	$SE(\widehat{\beta_0}) = 3.24916$	$Z_0 = -0.458$	0.646823	Accept
$\widehat{\beta_1} = -0.02765$	$SE(\widehat{\beta_1}) = 0.16490$	$Z_1 = -0.168$	0.866849	Accept
$\widehat{\beta_2} = -0.21158$	$SE(\widehat{\beta_2}) = 0.32732$	$Z_2 = -0.646$	0.518013	Accept
$\widehat{\beta_3} = 1.07337$	$SE(\widehat{\beta_3}) = 0.35316$	$Z_3 = 3.039$	0.002371	Reject
$\widehat{\beta_4} = 2.17757$	$SE(\widehat{\beta_4}) = 0.33478$	$Z_4 = 6.504$	7.8e-11	Reject
$\widehat{\beta_5} = 1.42000$	$SE(\widehat{\beta_5}) = 0.41033$	$Z_5 = 3.461$	0.000539	Reject

the fitted logistic regression equation (using the logit link function) is given as:

$$\widehat{\pi} = \frac{e^{-1.48870 - 0.02765 x_1 - 0.21158 x_2 + 1.07337 x_3 + 2.17757 x_4 + 1.42000 x_5}}{1 + e^{-1.48870 - 0.02765 x_1 - 0.21158 x_2 + 1.07337 x_3 + 2.17757 x_4 + 1.42000 x_5}}$$

PREDICTION

Case (1): p_1^* = First Quartile (Q1) = 0.343146

y	\hat{Y}_1	1	0	Total
1		114	40	154
0		70	28	98
Total		184	68	252

Case (2): p_2^* = Mean = 0.6111111

y_i	\hat{Y}_2	1	0	Total
1		91	63	154
0		47	51	98
Total		138	114	252

Case (3): p_3^* = Median (Q2) = 0.7839838

y	\hat{Y}_3	1	0	Total
1		79	75	154
0		40	58	98
Total		119	133	252

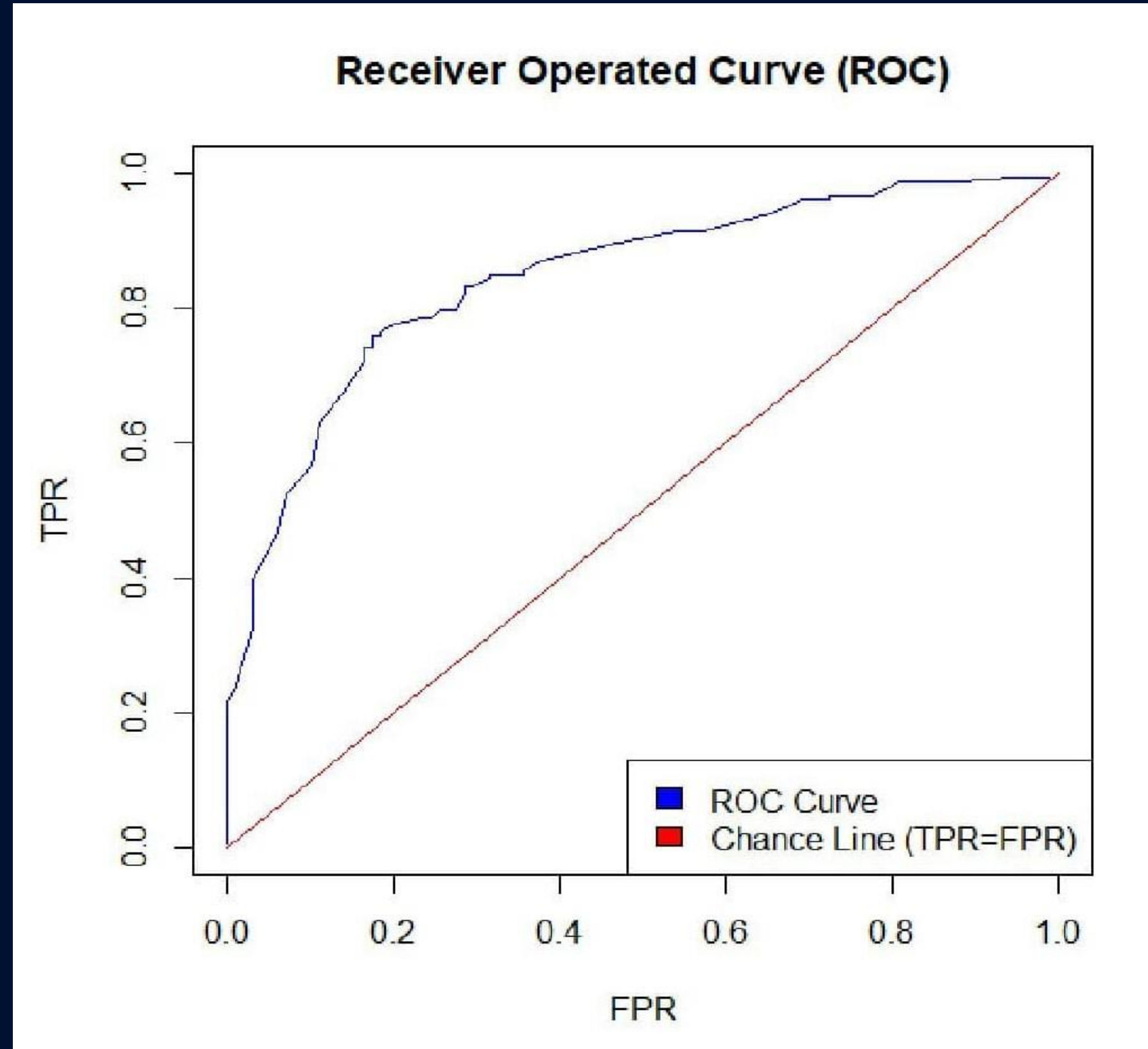
Case (4): p_4^* = Third Quartile (Q3) = 0.8297085

y	\hat{Y}_4	1	0	Total
1		39	115	154
0		23	75	98
Total		62	190	252

TABLE OF CALCULATIONS

Threshold	TPR	FPR	Total Probability of Misclassification
$p_1^* = 0.343146$	0.7402597	0.7142857	0.9740260
$p_2^* = 0.6111111$	0.5909091	0.4795918	0.8886827
$p_3^* = 0.7839838$	0.512987	0.408163	0.8951763
$p_4^* = 0.8297085$	0.2532468	0.2346939	0.9814471

ROC CURVE



CONCLUSION

- a) From the regression analysis,
 - the annual range of income,
 - previous and/or current cigarette use, and
 - having other friends who also vapeare significant factors influencing the development of vaping habits

- b) Vapes (Rs. 2000-7000 online) - only be afforded by higher income classes
(greater exposure through social media and peers who vape)

- c) 90% - have one or more friends who also vape
75-80% - first introduced to vaping through observing and/or copying their friends

THANK YOU