

Name: Neha Naidu

Project Title: Children's Safety in Cyberspace

Month: Feb 2025

1. Introduction:

Digital Exposure of Children:

In today's digital age, children are increasingly exposed to the internet through smartphones, tablets, online games, social media, and educational platforms from a very young age.

Rising Cyber Risks:

While digital platforms offer immense learning and entertainment opportunities, they also pose significant cyber risks such as cyberbullying, online predators, exposure to inappropriate content, privacy violations, and digital addiction.

Need for Awareness:

It becomes crucial for parents, educators, and society at large to understand the level of awareness about these risks and how well children are being protected online.

Objective of the Project:

This project aims to explore public perception and awareness related to children's safety in cyberspace. It investigates how well-informed individuals are about the threats and what preventive actions they take.

About the Dataset:

The analysis is based on primary data collected through an original online survey. The questionnaire received 55+ responses from a diverse audience, providing insights into their understanding, concerns, and practices regarding children's cybersecurity.

2. Data Collection:

How the Questionnaire Was Created

To gather real-world data on public awareness and perception of children's safety in cyberspace, a custom questionnaire was designed and distributed via Google Forms.

The questionnaire was structured to ensure clarity, accessibility, and relevance to both students and young adults, who are digitally aware and likely to interact with children or online platforms.

The form was shared through academic and social networks to ensure a diverse yet focused sample population.

Type of Questions

The survey comprised both quantitative and qualitative elements, structured into the following formats:

- Multiple Choice Questions (MCQs):
For demographic and awareness-based queries (e.g., *Gender, Awareness, Monitoring, Internet usage*).
- Rating Scale (Likert-type):
Respondents rated how safe they perceive children to be online (from 1 = Poor to 5 = Good).
- Checkbox (Multi-select) Questions:
Used for capturing multiple internet activities, safety practices, or parental guidance methods (e.g., *Online activities, Safety tips, Guidance techniques*).
- Open-ended / Opinion-based Fields:
Included to understand subjective opinions about online safety and parental control, such as in *CTUsage* and *talk* columns.

Sample Size

The survey collected **57 valid responses**, all of which were complete and usable for analysis.

Respondents included college students and young professionals from various backgrounds, offering a focused view into public perceptions of online safety for children.

Demographic Distribution

Age Group Breakdown

Age Group	Number of Respondents	Percentage %
18 - 25	52	91.2%
26 - 35	3	5.3%
36 - 50	1	1.8%
Above 50	1	1.8%

Insight: A dominant majority (91%) were between **18 and 25 years old**, reflecting a tech-savvy and digitally exposed population, ideal for analyzing opinions on cyberspace awareness.

Gender Distribution

Gender	Number of Respondents	Percentage %
Female	37	64.9 %
Male	20	35.1%

Insight: The survey had a higher proportion of **female respondents** (nearly 65%), which provides a broader scope for analyzing gender-wise differences in perception and awareness.

3. Dataset Overview

- Loading of Dataset and extracting first few lines

Code:

```
> safety<-read.csv(file.choose(),header=TRUE, sep=",")  
> head(safety[,c("Username","Name","AgeGroup","Gender","usage","Awareness","safe","rating")])  
      Username           Name AgeGroup Gender  
1 thoppil.sharon@gmail.com Sharon Thoppil 18 - 25 Female  
2 sharvpatil18@gmail.com Sharv Nitin Naik 18 - 25 Male  
3 vaishnum3004@gmail.com Vaishnavi More 18 - 25 Female  
4 shrutijhas2004@gmail.com Shruti jha 18 - 25 Female  
5 shrutiprasad0504@gmail.com Shruti Prasad Patil 18 - 25 Female  
6 bansodesamiksha00@gmail.com Samiksha Bansode 18 - 25 Female  
    usage Awareness safe rating  
1 Everyday Yes Not sure 3  
2 Everyday Yes No 2  
3 Everyday Yes No 1  
4 Everyday Yes Yes 3  
5 Everyday Yes No 4  
6 Everyday Yes Yes 4
```

- Structure of Dataset

Code:

```
> str(safety)  
'data.frame': 57 obs. of 22 variables:  
 $ Timestamp : chr "2024/12/09 11:18:21 PM GMT+5:30" "2024/12/09 11:20:17  
 PM GMT+5:30" "2024/12/09 11:29:05 PM GMT+5:30" "2024/12/09 11:36:25 PM GMT+5:  
 30" ...  
 $ Username  : chr "thoppil.sharon@gmail.com" "sharvpatil18@gmail.com" "va  
 ishum3004@gmail.com" "shrutijhas2004@gmail.com" ...  
 $ Name       : chr "Sharon Thoppil" "Sharv Nitin Naik" "Vaishnavi More"  
 "Shruti jha" ...  
 $ AgeGroup   : chr "18 - 25" "18 - 25" "18 - 25" "18 - 25" ...  
 $ Gender     : chr "Female" "Male" "Female" "Female" ...  
 $ Awareness   : chr "Yes" "Yes" "Yes" "Yes" ...  
 $ safe        : chr "Not sure" "No" "No" "Yes" ...  
 $ usage       : chr "Everyday" "Everyday" "Everyday" "Everyday" ...  
 $ CTUsage    : chr "It's safe as long as they're supervised" "It's safe as  
 long as they're supervised" "It's risky because there are many threats" "It's  
 risky because there are many threats" ...  
 $ activity    : chr "Study / Learning new things;Play games;Watch videos or  
 movies;Social media" "Study / Learning new things;Play games;Watch videos or  
 movies;Social media" "Play games;Social media" "Study / Learning new things"  
 ...  
 $ concerns    : chr "cyberbullying;Access to harmful content;spending too m  
 uch time online" "Cyberbullying;Access to harmful content;spending too much t  
 ime online" "spending too much time online" "Cyberbullying" ...  
 $ Monitor     : chr "Sometimes" "Sometimes" "Yes, all the time" "Yes, all t  
 he time" ...  
 $ talk         : chr "Yes, it's very important." "Yes, it's very important."  
 "Yes, it's very important." "Yes, it's very important." ...  
 $ Clearn       : chr "How to avoid harmful content;How to recognize online s  
 cams;How to deal with cyberbullying;How to manage screen time" "How to avoid  
 harmful content;How to recognize online scams;How to deal with cyberbullying;  
 How to manage screen time" "How to avoid harmful content;How to recognize on  
 line scams" "How to avoid harmful content;How to recognize online scams;How to  
 deal with cyberbullying" ...  
 $ onlinesafety: chr "Talking to family and friends" "Talking to family and  
 friends" "Workshops or classes" "Workshops or classes" ...  
 $ tips          : chr "Don't share personal information like address or phone  
 number;Be careful about talking to strangers online;Use '| __truncated__' 'Do  
 n't share personal information like address or phone number;Be careful about  
 talking to strangers online;Use '| __truncated__' 'Don't share personal inform  
 ation like address or phone number;Be careful about talking to strangers onli  
 ne" "Don't share personal information like address or phone number;Use strong  
 passwords" ...  
 $ safe_college: chr "yes" "yes" "yes" "yes" ...
```

```
$ guidance    : chr "Teach them not to talk to strangers;Limit their screen  
time;Use parental controls;Talk about online safety regularly" "Teach them no  
t to talk to strangers;Limit their screen time;Use parental controls;Talk abo  
ut online safety regularly" "Limit their screen time;Talk about online safety  
regularly" "Teach them not to talk to strangers;Limit their screen time;Use p  
arental controls" ...  
$ filters     : chr "Yes" "Yes" "Yes" "Yes" ...  
$ campaigns   : chr "Yes" "Yes" "Yes" "Yes" ...  
$ stayunsafe  : chr "Yes" "No" "Yes" "Yes" ...  
$ rating      : int 3 2 1 3 4 4 3 3 1 4 ...
```

- Statistical Summary of Dataset

Code:

```

> summary(safety)
Timestamp           Username          Name
Length:57          Length:57        Length:57
Class :character   Class :character Class :character
Mode  :character   Mode  :character Mode  :character

AgeGroup           Gender            Awareness
Length:57          Length:57        Length:57
Class :character   Class :character Class :character
Mode  :character   Mode  :character Mode  :character

safe               usage             CTUsage
Length:57          Length:57        Length:57
Class :character   Class :character Class :character
Mode  :character   Mode  :character Mode  :character

activity           concerns          Monitor
Length:57          Length:57        Length:57
Class :character   Class :character Class :character
Mode  :character   Mode  :character Mode  :character

talk               Clearn            onlineSafety
Length:57          Length:57        Length:57
Class :character   Class :character Class :character
Mode  :character   Mode  :character Mode  :character

tips               safe_college      guidance
Length:57          Length:57        Length:57
Class :character   Class :character Class :character
Mode  :character   Mode  :character Mode  :character

filters            campaigns         staySafe
Length:57          Length:57        Length:57
Class :character   Class :character Class :character
Mode  :character   Mode  :character Mode  :character

rating
Min.   :1.000
1st Qu.:2.000
Median :3.000
Mean   :2.825
3rd Qu.:4.000
Max.   :5.000

```

4. Data Cleaning

- Checking for Missing Values

Code:

```
> colSums(is.na(safety))
   Timestamp      Username        Name    AgeGroup
0             0            0          0           0
Gender     Awareness       safe    usage
0             0            0          0           0
CTUsage    activity    concerns  Monitor
0             0            0          0           0
talk      Cleanr onlineSafety     tips
0             0            0          0           0
safe_college guidance    filters campaigns
0             0            0          0           0
staysafe    rating  AgeGroupNum
0             0            0          0           0
` |
```

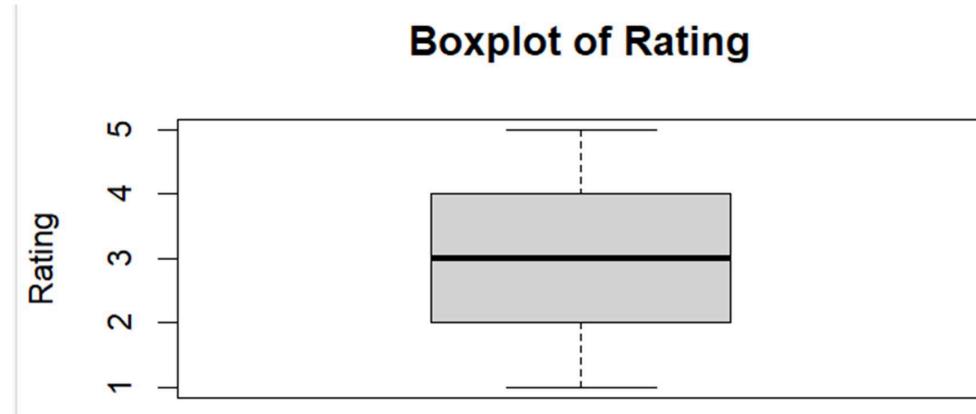
Insight: All columns returned **0 missing values**, indicating a **complete dataset**. No imputation or removal of rows was necessary.

- Outlier Detection

Code:

```
> # Boxplot for numerical columns like Rating
> boxplot(safety$rating, main = "Boxplot of Rating", ylab = "Rating")
```

Graph:



Conclusion:

- No Outliers Present - The boxplot does not show any points outside the whiskers, indicating that there are no extreme values (outliers) in the data.
- Since no outliers exist, there is no need for outlier treatment, such as removal, transformation, or imputation.

5. Exploratory Data Analysis

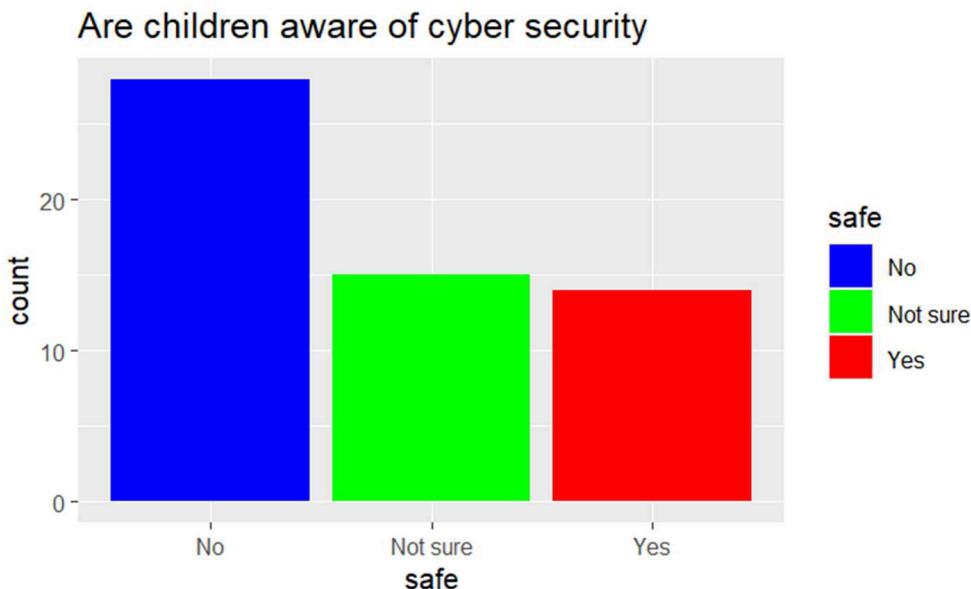
a) Univariate Analysis

1. Do children have an understanding of how to stay safe online?

Code:

```
> ggplot(safety, aes(x = safe, fill = safe)) +  
+   geom_bar() +  
+   ggtitle("Are children aware of cyber security") +  
+   scale_fill_manual(values = c("blue", "green", "red"))
```

Graph:



Conclusion:

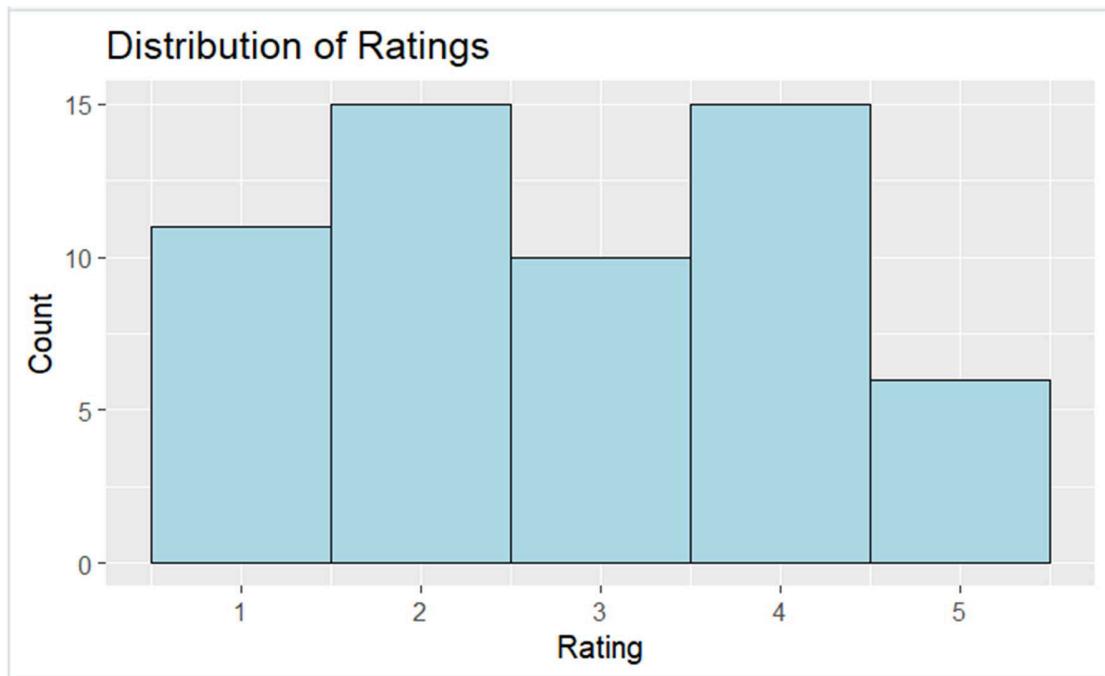
- The graph suggests that a significant portion of respondents believe that children are not aware of cybersecurity, as indicated by the highest count in the "No" category.
- A smaller number of respondents are unsure, while an even smaller group believes that children are aware of cybersecurity.
- This indicates a general concern that children may not be sufficiently educated about online safety, highlighting the need for increased awareness and education on cybersecurity for young users.

2. How well-protected are children in cyberspace, according to public perception?

Code:

```
> ggplot(safety, aes(x = rating)) +  
+   geom_histogram(binwidth = 1, fill = "lightblue", color = "black") +  
+   ggttitle("Distribution of Ratings") +  
+   xlab("Rating") +  
+   ylab("Count")
```

Graph:



1 – poor 2 – weak 3 – marginal 4 – adequate 5 – good

Conclusion:

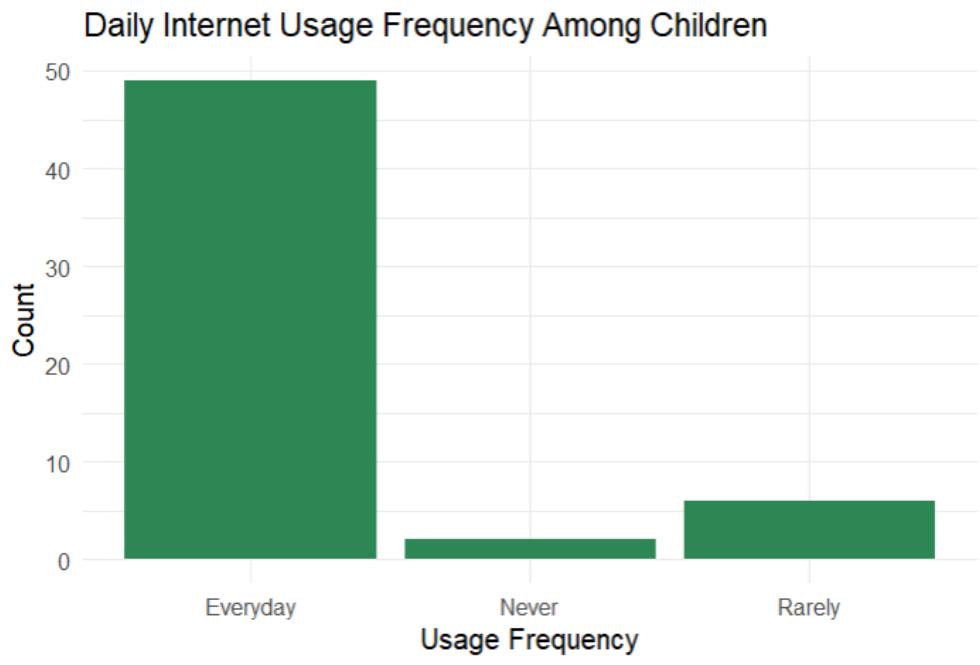
- The graph indicates that the majority of respondents (30 individuals) have rated children's protection in cyberspace with 2 and 4 stars. This suggests a divided perception, where 15 people (50% of this majority) believe children are weakly protected, while the other 15 (50%) consider them adequately protected.
- Additionally, more than 10 respondents perceive children as poorly protected, 10 believe they are marginally protected, and fewer than 10 think children are well-protected.
- These statistics highlight concerns regarding children's online safety, with a notable portion of respondents indicating insufficient protection.

3. In your opinion, how often do children use the internet?

Code:

```
ggplot(df, aes(x = usage)) +geom_bar(fill = "seagreen") +  
  labs(title = "Daily Internet Usage Frequency Among Children",  
    x = "Usage Frequency", y = "Count") +theme_minimal()
```

Graph:



Conclusion:

- Out of **57 respondents**, **49** ($\approx 86\%$) reported that children use the internet **every day**, indicating a high level of consistent digital engagement.
- Only **2 respondents** said children **rarely** use the internet, and **1 respondent** said **never**, showing that limited access is uncommon.
- The overwhelming frequency of daily internet use suggests **increased exposure to both opportunities and risks online**, reinforcing the need for **cyber safety education, monitoring tools, and digital boundaries**.

b) Bivariate Analysis

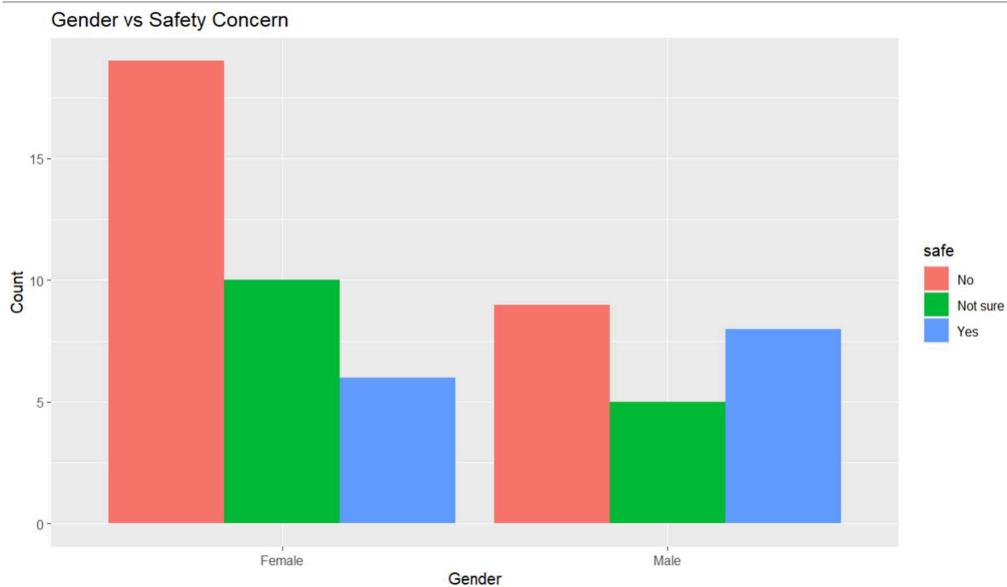
- Categorical vs Categorical

1. Genderwise opinion on how well aware children are about cybersecurity.

Code:

```
> library(ggplot2)
> ggplot(safety, aes(x = Gender, fill = safe)) +
+   geom_bar(position = "dodge") +
+   labs(title = "Gender vs Safety Concern", x = "Gender", y = "Count")
```

Graph:



Conclusion:

- Out of a total of 57 respondents, the majority of females (less than 20) believe that children are not safe in cyberspace, while 10 are uncertain, and fewer than 10 consider children safe.
- Among males, fewer than 10 share the concern that children are not safe, while 5 are unsure, and approximately 7–8 believe children are safe.
- These statistics suggest that a higher proportion of females perceive children as unsafe online compared to males and proportion of uncertainty is relatively similar between genders, but slightly more males believe that children are adequately protected.
- Additionally, the number of individuals who believe children are safe remains relatively low across both genders. This distribution highlights a prevailing concern regarding children's cybersecurity.

2. What percentage of females and males are aware of online safety issues for children?

Code:

```
> table(safety$Gender, safety$Awareness)
```

	No	Yes
Female	3	32
Male	3	19

Conclusion:

- The table represents the relationship between gender and awareness of children's exposure to cyber threats such as cyberbullying and online scams.
- Among females, 32 are aware of these issues, while only 3 are not and among males, 19 are aware, while 3 are not.
- This shows that the majority of both females (91.4%) and males (86.4%) are aware of cyber threats affecting children, indicating a high level of awareness across genders.
- There is a strong awareness of children's exposure to cyber risks across both genders, with females showing slightly higher awareness than males.
- This suggests that awareness campaigns or discussions on cyber safety are reaching a broad audience, though targeted efforts could be made to ensure the remaining individuals become informed.

- **Categorical vs Numerical**

1. What is the relationship between gender and the perception of children's protection in cyberspace?

Code:

```
> chisq.test(table(safety$Gender,safety$rating))

Pearson's Chi-squared test

data: table(safety$Gender, safety$rating)
X-squared = 9.3966, df = 4, p-value = 0.05192
```

Conclusion:

- The p-value (0.05192) is slightly above the typical significance level of 0.05, indicating that there is not enough evidence to reject the null hypothesis at the 5% significance level.
- Therefore, we conclude that there is no statistically significant association between gender and the perception of children's protection in cyberspace at the 0.05 level.
- However, since the p-value is very close to 0.05, this suggests that there might be a weak association between the two variables.

- Numerical vs Numerical

1. Are there significant differences in how different age groups rate the protection of children in cyberspace?

Code:

```
> anova_result <- aov(rating ~ AgeGroup, data = safety)
> summary(anova_result)
Df Sum Sq Mean Sq F value Pr(>F)
AgeGroup      2   1.45   0.7228   0.412  0.665
Residuals    54  94.80   1.7556
```

Conclusion:

- From one-way ANOVA test we get following results: F-value: 0.412 and p-value: 0.665.
- Since the p-value (0.665) is greater than the standard significance level of 0.05, we fail to reject the null hypothesis which means there is no statistically significant difference in the ratings of children's protection in cyberspace across the different age groups.
- The age group does not appear to significantly influence how protected individuals feel children are in cyberspace based on the current data.

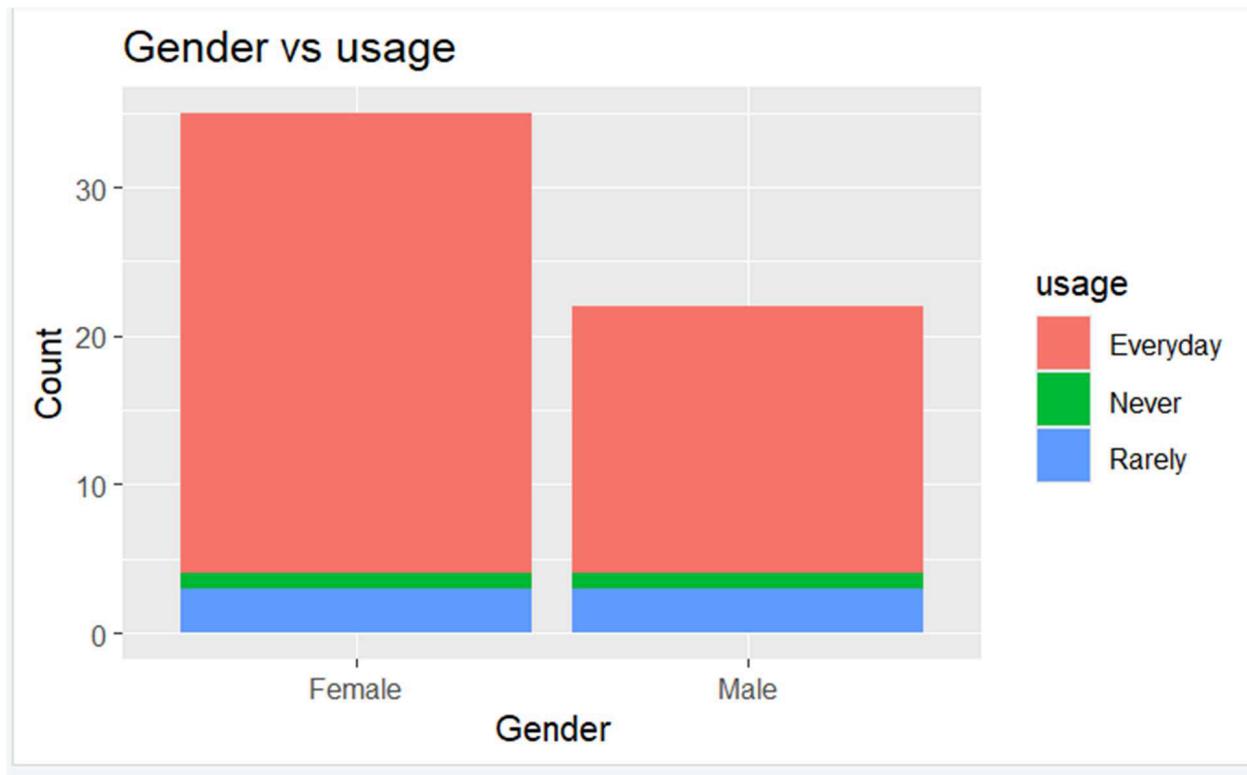
6. Data Visualization

1. What is the most common usage pattern among males and females?

Code:

```
> ggplot(safety, aes(x = Gender, fill = usage)) +  
+   geom_bar(position = "stack") +  
+   labs(title = "Gender vs usage", x = "Gender", y = "Count")  
~
```

Graph:



Conclusion:

- The number of females who use internet Everyday is notably higher than the number of males.
- The "Rarely" usage category shows a higher count for males compared to females.
- The "Never" usage category is relatively low and fairly similar for both genders.

2. Which gender group has a higher proportion of awareness regarding online safety?

Code:

```
> library(ggplot2)
> ggplot(safety, aes(x = Gender, y = Awareness)) +
+   geom_tile(aes(fill = rating)) +
+   scale_fill_gradient(low = "blue", high = "red")
```

Graph:



Conclusion:

1. Gender Differences in Awareness:

- Females show a uniform rating distribution for both "Yes" and "No" awareness groups.
- Males show more variation, with a significant portion of those who are aware (Yes) giving lower ratings (blue color indicating lower ratings).

2. Ratings Distribution:

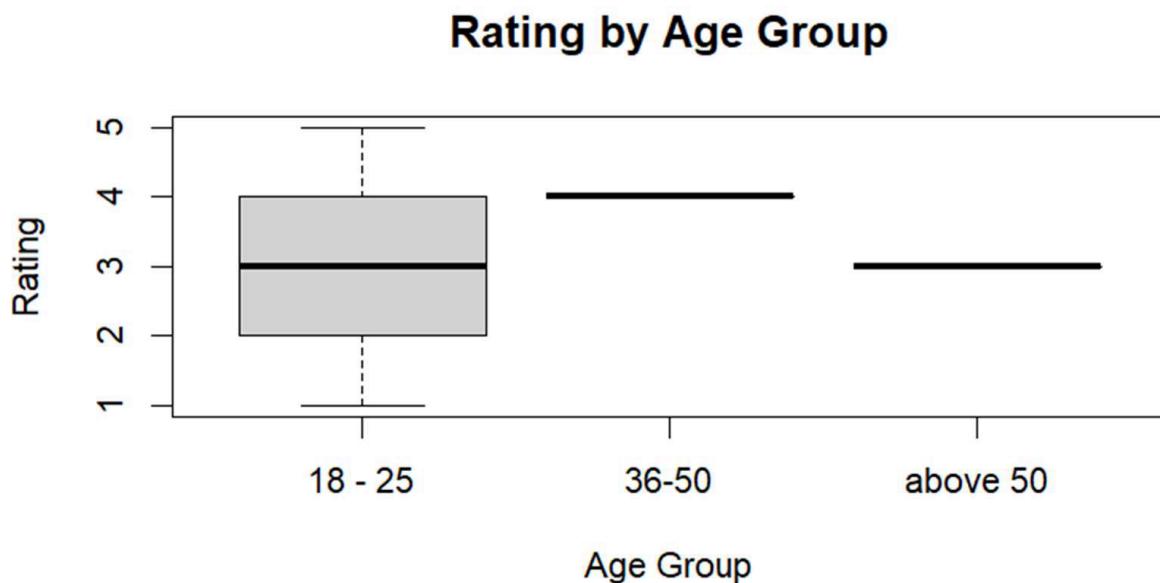
- The blue section in males indicates a lower rating (1-2) for awareness, suggesting that even those who are aware may not perceive strong online safety.
- Females seem to have a more consistent middle-to-high rating (pinkish-red color).

3. How does the perception of children's safety in cyberspace vary across different age groups?

Code:

```
> boxplot(rating ~ AgeGroup, data = safety, main = "Rating by Age Group", xlab = "Age Group", ylab = "Rating")
```

Graph:



Conclusions:

- 18-25 Age Group:
The median rating is around 3 and whiskers extend from approximately 1 to 5, meaning responses vary widely.
- 36-50 and Above 50 Age Groups:
Minimal variation in ratings as ratings appear more consistent and clustered around a single value.
- The above 50 group shows ratings mostly around 2.5, indicating a lower perceived safety.

From this we can conclude the following:

- Younger participants (18-25) have a more diverse perception of children's safety in cyberspace.
- Older participants (36-50, above 50) have more stable ratings, suggesting a stronger consensus.
- No visible outliers, meaning there are no extreme deviations in ratings.