```
Shriya Parab
2021700044

# Read the CSV file
data <- read.csv("/content/Dataset_CyberCrime_Sean.csv")
data

# Install ggplot2 if not already installed
install.packages("ggplot2")

# Load libraries
library(ggplot2)
```

	City	Personal.Revenge	Anger	Fraud	Extortion
1	Agra	5	0	19	0
2	Allahabad	0	0	222	11
3	Amritsar	2	0	5	0
4	Asansol	6	1	3	0
5	Aurangabad	5	2	51	0
6	Bhopal	0	0	4	7
7	Chandigarh City	0	0	19	3
8	Dhanbad	2	0	29	0
9	Durg-Bhilainagar	0	0	0	0
10	Faridabad	0	0	9	0
11	Gwalior	0	2	24	1
12	Jabalpur	0	0	0	0
13	Jamshedpur	0	0	47	0
14	Jodhpur	0	0	59	0
15	Kannur	3	1	1	0
16	Kollam	0	0	7	0
17	Kota	0	3	6	5
18	Ludhiana	0	0	6	4
19	Madurai	0	0	0	0

20	Malappuram	2	1	2	1
21	Meerut	0	0	0	0
22	Nasik	0	0	21	0
23	Raipur	0	0	3	0
24	Rajkot	0	0	3	0
25	Ranchi	0	0	175	0
26	Srinagar	1	0	6	1
27	Thiruvananthapuram	7	2	23	0
28	Thrissur	0	0	3	0
29	Tiruchirapalli	1	0	0	0
30	Vadodara	0	0	11	0
⋮	⋮	⋮	⋮	⋮	⋮
162	Jharkhand	4	4	1069	14
163	Karnataka	147	13	9680	74
164	Kerala	44	34	96	21
165	Madhya Pradesh	7	6	292	13
166	Maharashtra	36	105	3413	45
167	Manipur	0	2	40	0
168	Meghalaya	6	10	81	7
169	Mizoram	0	0	3	0
170	Nagaland	0	0	5	0
171	Odisha	1	33	1380	175
172	Punjab	4	19	164	29
173	Rajasthan	22	10	641	42
174	Sikkim	0	0	0	0
175	Tamil Nadu	83	57	134	112

176	Telangana	96	24	4436	115
177	Tripura	14	1	11	0
178	Uttar Pradesh	78	210	4674	1055
179	Uttarakhand	11	5	98	33
180	West Bengal	66	8	72	12
181	Total State(s)	1463	814	30075	2411
182	A & N Islands	0	0	0	1
183	Chandigarh	0	0	7	1
184	D & N Haveli and Daman & Diu	0	0	0	0
185	Delhi	2	4	23	15
186	Jammu & Kashmir	3	4	33	9
187	Ladakh	0	0	0	0
188	Lakshadweep	2	0	0	0
189	Puducherry	0	0	4	3
190	Total UT(s)	7	8	67	29
191	Total All India	1470	822	30142	2440

	Causing.Disrepute	Prank	Sexual.Exploitation	Disrupt.Public.Service
1	0	0	0	0
2	8	0	0	0
3	0	0	2	0
4	0	0	0	0
5	0	0	21	0
6	2	0	1	1
7	0	7	0	0
8	0	0	0	0
9	10	0	0	0
10	0	0	0	0
11	9	0	0	0
12	0	0	0	0
13	0	0	0	0
14	4	0	6	0
15	0	0	1	0
16	0	0	5	0
17	5	3	3	0
18	0	0	13	0

```
19   0                0      0                 0
20   0                0      0                 0
21  53                0      0                 0
22   8                0     12                 0
23   8                0      9                 0
24   0                0      0                 0
25   0                0      0                 0
26   5                0      0                 0
27   2                1      2                 0
28   0                0      2                 0
29   0                0      0                 0
30   1                1      0                 0
⋮   ⋮                ⋮    ⋮                 ⋮
162    2                0     13                 7
163  368                0    191                 1
164   58               10    138                 0
165   66                2    119                 0
166   76               32    612                 2
167    3                0     10                 0
168    9                0      9                 0
169    2                3      1                 0
170    1                1      0                 0
171    0                0    239                 0
172   19                3     58                 2
173   73               11     67                 0
174    0                0      0                 0
175   43                7    192                16
176    3                0     85                 1
177    2                0      3                 0
178  547               87    560                35
179    6                0     44                 0
180    3                3     44                 0
181 1678              252   3249                90
182    0                0      2                 0
183    0                0      7                 0
184    0                0      3                 0
185    0                0     20                 0
186   28                2     12                 1
187    0                0      0                 1
188    0                0      0                 0
189    0                0      0                 0
190   28                2     44                 2
191 1706              254   3293                92
    Sale.purchase.illegal.drugs Developing.own.business
Spreading.Piracy
1   0                           0                      0

2   0                           0                      0
```

3	0	0	0
4	0	0	0
5	0	0	0
6	0	0	0
7	0	0	0
8	0	0	0
9	0	0	0
10	0	0	0
11	0	0	0
12	0	0	0
13	0	0	0
14	0	0	0
15	0	0	0
16	0	0	0
17	0	0	0
18	0	0	0
19	0	0	0
20	0	0	0
21	0	0	0
22	0	0	0
23	0	0	0
24	0	0	0
25	0	0	0
26	0	0	0
27	0	0	0
28	0	0	0

29	0	0	0
30	0	0	0
⋮	⋮	⋮	⋮
162	13	0	0
163	2	17	4
164	0	0	1
165	0	7	0
166	0	14	0
167	0	0	0
168	0	0	0
169	0	0	0
170	0	0	0
171	0	0	0
172	0	1	3
173	0	17	0
174	0	0	0
175	0	0	19
176	0	0	0
177	0	0	1
178	0	76	36
179	4	2	0
180	0	1	0
181	21	203	75
182	0	0	0
183	0	0	0
184	0	0	0
185	0	0	0

186	0	7	0
187	0	0	0
188	0	0	0
189	0	0	0
190	0	7	0
191	21	210	75

	Psycho.or.Pervert	Steal.Information	Abetment.to.Suicide	Others	Total
1	0	0	0	46	70
2	0	0	0	0	241
3	0	0	0	0	9
4	0	0	0	11	21
5	0	0	0	0	82
6	0	0	0	0	16
7	0	0	0	1	30
8	0	0	0	0	31
9	0	0	0	0	10
10	0	0	0	8	17
11	0	0	0	0	41
12	0	0	0	0	0
13	0	0	0	0	47
14	0	0	0	42	111
15	0	0	0	1	10
16	0	0	0	0	12
17	0	0	0	0	26
18	0	0	0	0	24
19	0	0	0	0	0

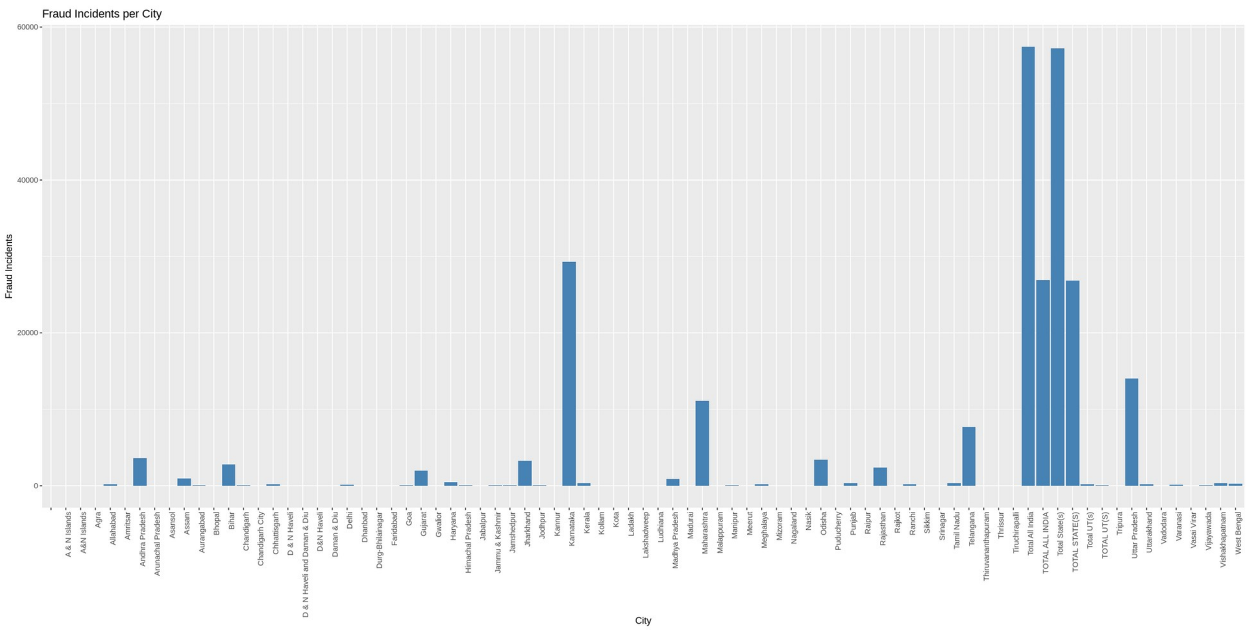
20	0	0	0	0	7
21	0	0	0	11	64
22	0	0	0	0	41
23	0	0	0	2	23
24	0	0	0	4	7
25	0	0	0	0	175
26	0	0	0	1	15
27	0	0	0	4	47
28	0	0	0	0	5
29	0	0	0	4	6
30	0	0	0	0	15
⋮	⋮	⋮	⋮	⋮	⋮
162	0	49	0	22	1204
163	0	0	0	223	10741
164	0	0	0	14	426
165	0	0	0	184	699
166	0	0	0	1149	5496
167	0	0	0	14	79
168	0	1	0	17	142
169	0	0	0	0	13
170	0	0	0	1	8
171	0	0	0	103	1931
172	0	0	0	67	378
173	0	0	0	463	1354
174	0	0	0	0	0

```
175 0                  0                 0                        11
782
176 0                  0                 0                       256
5024
177 0                  0                 0                         1
34
178 0                  5                 0                      3483
11097
179 0                  0                 0                        39
243
180 0                  0                 0                       502
712
181 0                 62                 0                      8688
49708
182 0                  0                 0                         2
5
183 0                  0                 0                         2
17
184 0                  0                 0                         0
3
185 0                  0                 0                       104
168
186 0                  0                 0                        14
120
187 0                  0                 0                         0
1
188 0                  0                 0                         1
3
189 0                  0                 0                         3
10
190 0                  0                 0                       126
327
191 0                 62                 0                      8814
50035

Installing package into ‘/usr/local/lib/R/site-library’
(as ‘lib’ is unspecified)
```

```
# Bar chart of Fraud incidents per city
ggplot(data, aes(x = City, y = Fraud)) +
  geom_bar(stat = "identity", fill = "steelblue") +
  theme(axis.text.x = element_text(angle = 90, hjust = 1)) +
  labs(title = "Fraud Incidents per City", x = "City", y = "Fraud
Incidents")
```

```
Warning message:
“Removed 1 row containing missing values or values outside the scale
range
(`geom_bar()`).”
```



Obseravtions:

High Fraud Incidents: The city with the highest number of fraud incidents is "Andhra Pradesh," followed by "Telangana."

Low Fraud Incidents: Many cities have very low or no reported fraud incidents. These include "Arunachal Pradesh," "Chhattisgarh," "Goa," "Himachal Pradesh," "Jharkhand," "Meghalaya," "Nagaland," "Sikkim," and "Tripura."

Clustered Fraud Incidents: There seems to be a cluster of cities with relatively high fraud incidents in the central and southern regions of India.

Outlier: "Visakhapatnam" stands out as a city with a significantly higher number of fraud incidents compared to other cities in the southern region.

```r
# Remove rows where the 'Fraud' incidents are zero or missing (NA)
cleaned_fraud <- data[!is.na(data$Fraud) & data$Fraud > 0, ]

# Check if the data is clean
print(cleaned_fraud$Fraud)
```

```
  [1]    19   222     5     3    51     4    19    29     9    24
47    59
 [13]     1     7     6     6     2    21     3     3   175     6
23     3
 [25]    11   111    16    75   358   537    48   397    47     8
305    43
 [37]     8    15   460  2764    77   155  2171    35    15     7
531    53
 [49]   331    51   529     3  3450    54    45 12139    24    50
74 12213
 [61]   733     2   389   351    23    11   401   137    18    20
```

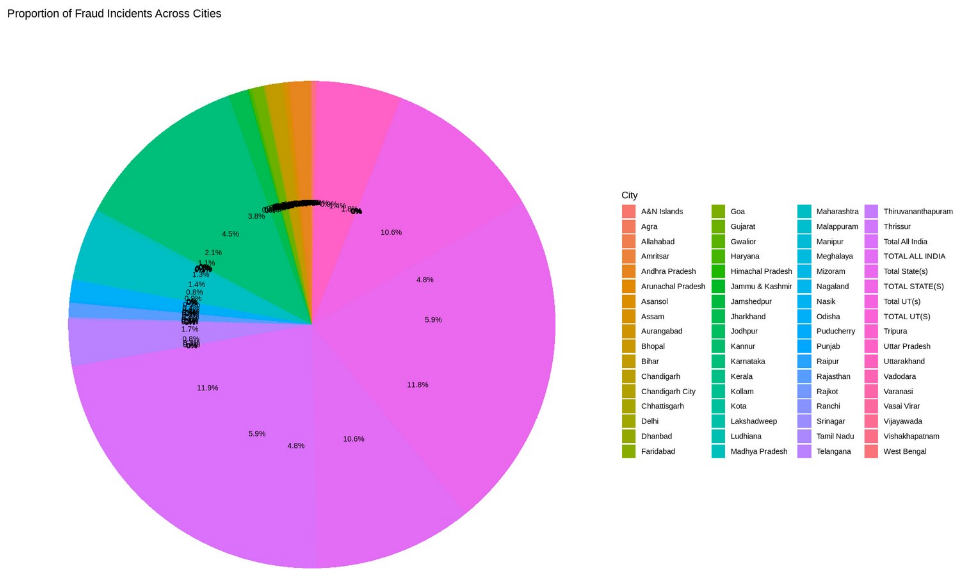
```
783  5441
 [73]    93   230  1998    14    35   506    48   499    55   732
8  2351
 [85]    46    68 14992     3    19    36     1    59 15051  1211
243   844
 [97]    47     9   363   151    15    27   964 11381    67   178
3551    50
[109]  1000    81   938    73  2013     2  3549    28    68 26853
7    31
[121]    38 26891  1149    26   242  1218    75    25   875   157
19  1069
[133]  9680    96   292  3413    40    81     3     5  1380   164
641   134
[145]  4436    11  4674    98    72 30075     7    23    33     4
67 30142

# Load ggplot2
library(ggplot2)

# Ensure city is a factor
cleaned_fraud$City <- factor(cleaned_fraud$City)

# Calculate percentages for the labels
cleaned_fraud$percentage <- (cleaned_fraud$Fraud /
sum(cleaned_fraud$Fraud)) * 100

# Create the pie chart
ggplot(cleaned_fraud, aes(x = "", y = Fraud, fill = City)) +
  geom_bar(stat = "identity", width = 1) +
  coord_polar("y") +
  labs(title = "Proportion of Fraud Incidents Across Cities") +
  theme_void() +  # Remove axes and gridlines
  theme(legend.position = "right") +  # Move legend to the right for
better clarity
  geom_text(aes(label = paste0(round(percentage, 1), "%")),
            position = position_stack(vjust = 0.5),  # Position labels
in the middle of the slices
            size = 3)  # Adjust text size to make it fit better
```



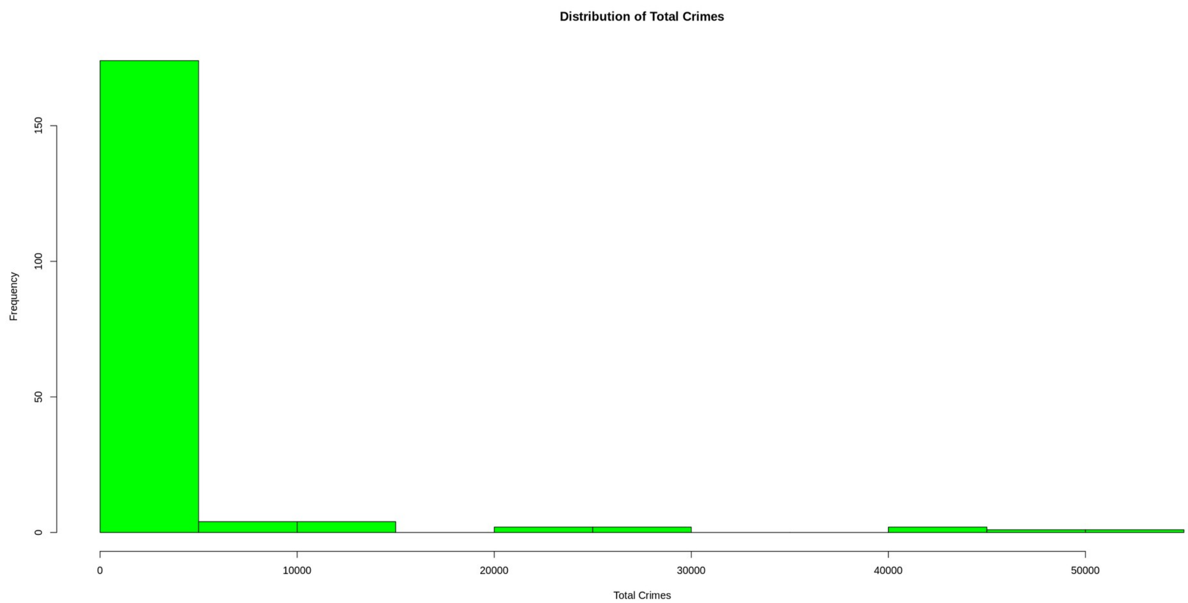
Observations:

Dominant Cities: The cities with the highest proportions of fraud incidents are "Maharashtra" and "Telangana," accounting for 10.6% and 10.2% respectively.

Large Share: "Tamil Nadu" also has a significant share of fraud incidents, contributing 9.6% to the total.

Smaller Shares: The remaining cities have relatively smaller proportions of fraud incidents, ranging from 0.1% to 3.7%.

```
# Histogram for 'Total' crimes across all cities
hist(data$Total,
     col="green",
     main="Distribution of Total Crimes",
     xlab="Total Crimes",
     ylab="Frequency")
```



Observations:

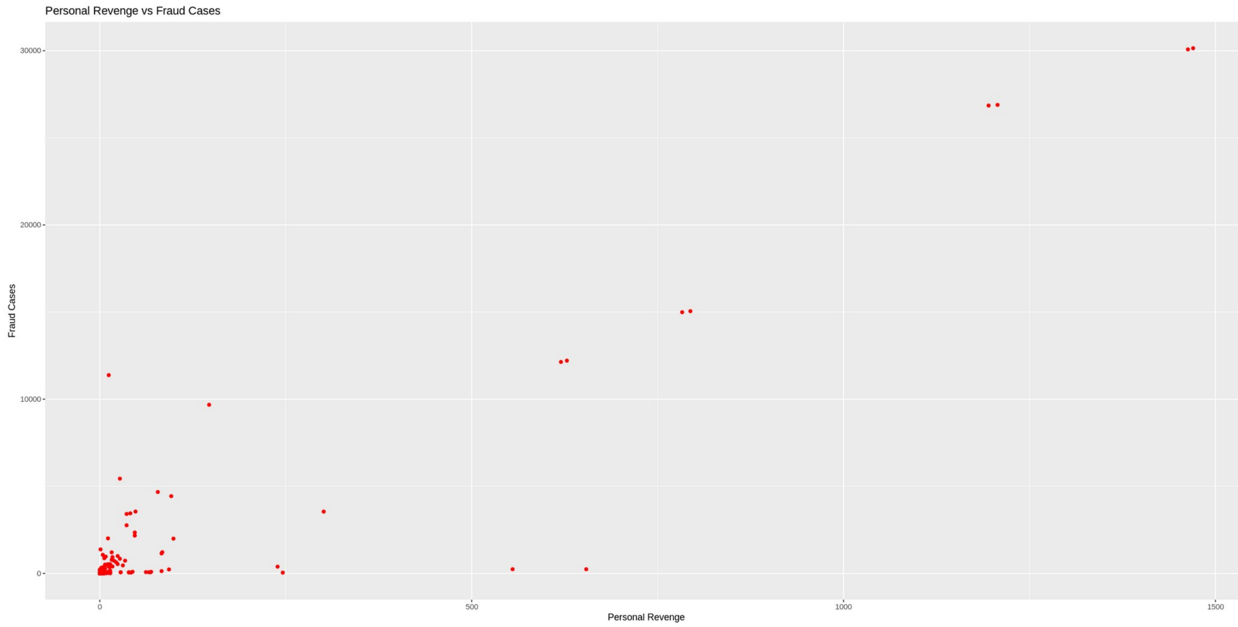
Skewness: The distribution is heavily skewed to the right, indicating that there are a few cities with a very high number of total crimes, while the majority of cities have relatively low crime rates.

Clustering: There seems to be a cluster of cities with total crimes between 5,000 and 10,000.

Outlier: The city with the highest number of total crimes is a clear outlier, falling far to the right of the main cluster.

```
# Scatter plot of 'Personal Revenge' vs 'Fraud'
ggplot(data, aes(x=`Personal.Revenge`, y=Fraud)) +
  geom_point(color="red") +
  labs(title="Personal Revenge vs Fraud Cases", x="Personal Revenge",
y="Fraud Cases")

Warning message:
“Removed 1 row containing missing values or values outside the scale
range
(`geom_point()`).”
```



Observations:

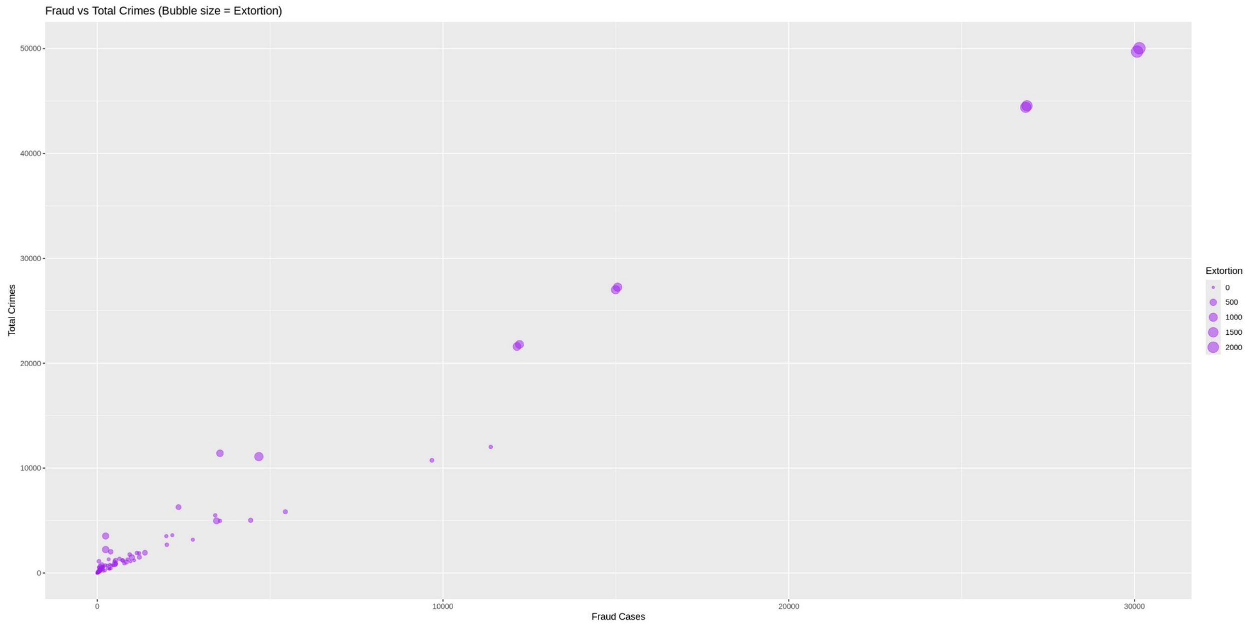
No Clear Correlation: There doesn't appear to be a strong linear relationship between the two types of cases. The points are scattered across the plot, indicating that there's no clear pattern of one type of case increasing or decreasing with another.

Outliers: There are a few outliers, particularly in the upper right corner, suggesting that some cities have exceptionally high numbers of both personal revenge cases and fraud cases.

Clustering: There might be a slight clustering of points in the lower left corner, indicating that some cities have relatively low numbers of both types of cases.

```
# Bubble plot: Fraud vs Total, with bubble size representing Extortion
ggplot(data, aes(x=Fraud, y=Total, size=Extortion)) +
  geom_point(alpha=0.5, color="purple") +
  labs(title="Fraud vs Total Crimes (Bubble size = Extortion)",
x="Fraud Cases", y="Total Crimes")

Warning message:
“Removed 1 row containing missing values or values outside the scale
range
(`geom_point()`).”
```



Observations:

Positive Correlation: There seems to be a positive correlation between fraud cases and total crimes. As the number of fraud cases increases, the number of total crimes tends to increase as well.

Extortion as a Factor: The size of the bubbles (representing extortion) suggests that extortion might be a contributing factor to both fraud cases and total crimes. Larger bubbles indicate higher levels of extortion, and these points tend to be located in the upper right corner of the plot, where both fraud cases and total crimes are higher.

Outliers: There are a few outliers, particularly in the upper right corner, indicating cities with exceptionally high numbers of fraud cases, total crimes, and extortion.