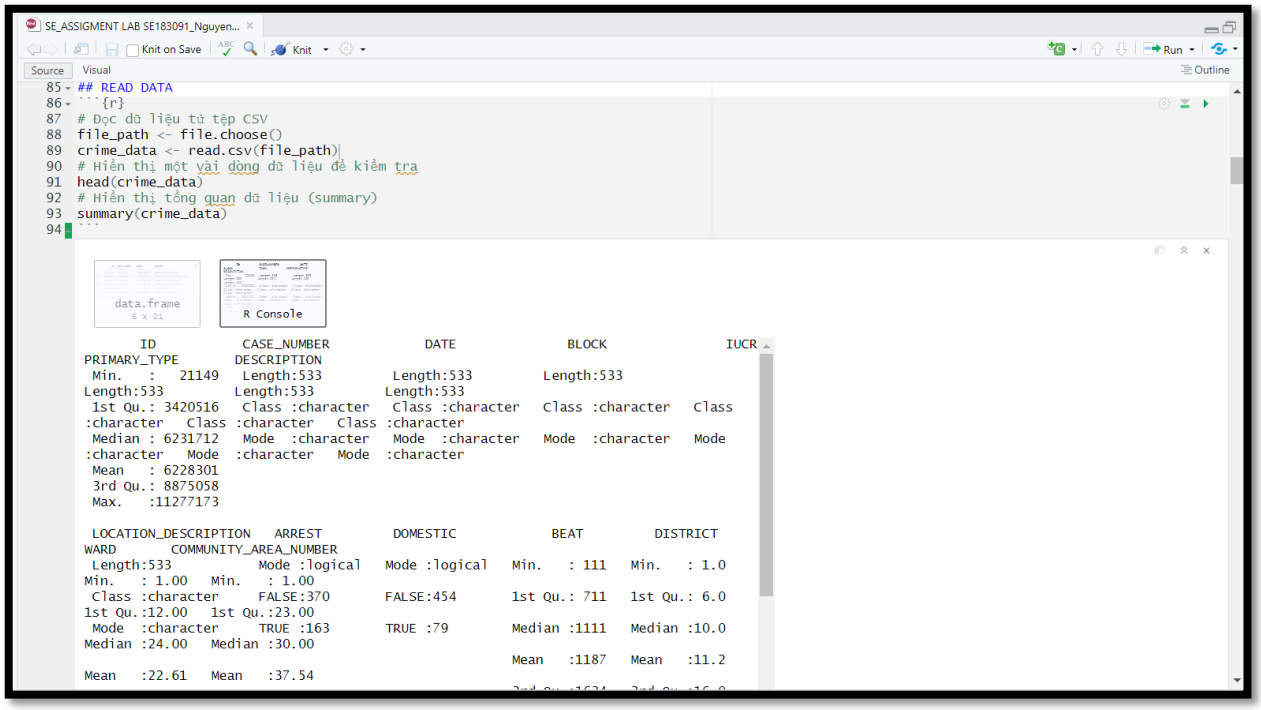
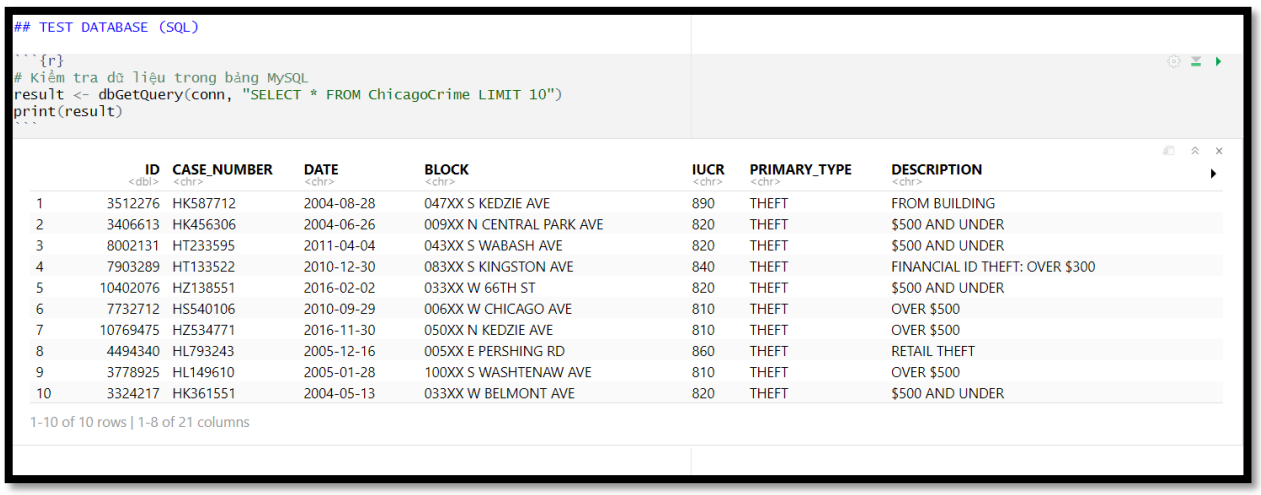
SE183091_Nguyễn Thanh Hòa



Tập dữ liệu có 533 mục với 21 cột, mô tả các vụ phạm tội ở Chicago. Các cột chính bao gồm:
- **ID**: Mã định danh duy nhất cho mỗi vụ.
- **CASE_NUMBER**: Số hồ sơ chính thức.
- **DATE**: Ngày xảy ra vụ việc.
- **PRIMARY_TYPE**: Loại tội phạm chính (ví dụ: trộm cắp).
- **DESCRIPTION**: Mô tả chi tiết về tội phạm.
- **ARREST**: Giá trị boolean chỉ ra liệu có bắt giữ hay không.
- **DOMESTIC**: Giá trị boolean chỉ ra tội phạm có liên quan đến bạo lực gia đình hay không.
- **LOCATION_DESCRIPTION**: Mô tả địa điểm xảy ra tội phạm (ví dụ: nhà ở, hẻm).
- **LATITUDE/LONGITUDE**: Tọa độ địa lý của nơi xảy ra tội phạm.



	ID <dbl>	CASE_NUMBER <chr>	DATE <chr>	BLOCK <chr>	IUCR <chr>	PRIMARY_TYPE <chr>	DESCRIPTION <chr>
1	3512276	HK587712	2004-08-28	047XX S KEDZIE AVE	890	THEFT	FROM BUILDING
2	3406613	HK456306	2004-06-26	009XX N CENTRAL PARK AVE	820	THEFT	$500 AND UNDER
3	8002131	HT233595	2011-04-04	043XX S WABASH AVE	820	THEFT	$500 AND UNDER
4	7903289	HT133522	2010-12-30	083XX S KINGSTON AVE	840	THEFT	FINANCIAL ID THEFT: OVER $300
5	10402076	HZ138551	2016-02-02	033XX W 66TH ST	820	THEFT	$500 AND UNDER
6	7732712	HS540106	2010-09-29	006XX W CHICAGO AVE	810	THEFT	OVER $500
7	10769475	HZ534771	2016-11-30	050XX N KEDZIE AVE	810	THEFT	OVER $500
8	4494340	HL793243	2005-12-16	005XX E PERSHING RD	860	THEFT	RETAIL THEFT
9	3778925	HL149610	2005-01-28	100XX S WASHTENAW AVE	810	THEFT	OVER $500
10	3324217	HK361551	2004-05-13	033XX W BELMONT AVE	820	THEFT	$500 AND UNDER

1-10 of 10 rows | 1-8 of 21 columns

Problem 1
Total number of cases

```{r}
query <- "
SELECT COUNT(*) AS total_cases
FROM ChicagoCrime
"
# Thực thi truy vấn
result <- dbGetQuery(conn, query)
# In kết quả
print(result)
```

	total_cases <chr>
1	533

1 row

Problem 2
Total number of cases by crime type

```{r}
query <- "
SELECT PRIMARY_TYPE AS crime_type, COUNT(*) AS total_cases
FROM ChicagoCrime
GROUP BY PRIMARY_TYPE
"
# Thực thi truy vấn
result <- dbGetQuery(conn, query)
# In kết quả
print(result)
```

	crime_type <chr>	total_cases <chr>
1	ARSON	2
2	ASSAULT	32
3	BATTERY	92
4	BURGLARY	30
5	CONCEALED CARRY LICENSE VIOLATION	1
6	CRIM SEXUAL ASSAULT	3
7	CRIMINAL DAMAGE	58
8	CRIMINAL TRESPASS	15
9	DECEPTIVE PRACTICE	20
10	DOMESTIC VIOLENCE	1

1-10 of 35 rows Previous 1 2 3 4 Next

Problem 3 **Total number of cases by year**

```{r}
query <- "
SELECT YEAR, COUNT(*) AS total_cases
FROM ChicagoCrime
GROUP BY YEAR
"
# Thực thi truy vấn
result <- dbGetQuery(conn, query)
# In kết quả
print(result)
```

	YEAR <dbl>	total_cases <chr>
1	2001	36
2	2002	37
3	2003	42
4	2004	33
5	2005	44
6	2006	30
7	2007	36
8	2008	21
9	2009	43
10	2010	25

1-10 of 18 rows Previous 1 2 Next

Problem 4 **Number of cases without arrests (Non-Arrest)**

```{r}
query <- "
SELECT COUNT(PRIMARY_TYPE) AS total_cases_non_arrest
FROM ChicagoCrime
WHERE ARREST = 'FALSE'
"
# Thực thi truy vấn
result <- dbGetQuery(conn, query)
# In kết quả
print(result)
```

	total_cases_non_arrest <chr>
1	370

1 row

Problem 5 **Total number of cases by community area**

```{r}
query <- "
SELECT COMMUNITY_AREA_NUMBER, COUNT(*) AS total_cases
FROM ChicagoCrime
GROUP BY COMMUNITY_AREA_NUMBER
ORDER BY COMMUNITY_AREA_NUMBER;
"
# Thực thi truy vấn
result <- dbGetQuery(conn, query)
# In kết quả
print(result)
```

	COMMUNITY_AREA_NUMBER <dbl>	total_cases <chr>
1	*NA*	43
2	1	6
3	2	7
4	3	4
5	4	3
6	5	4
7	6	11
8	7	5
9	8	15
10	10	3

1-10 of 71 rows

Previous 1 2 3 4 5 6 … 8 Next

Problem 6 Number of cases with missing location information

```{r}
query <- "
SELECT COUNT(*) AS total_cases
FROM ChicagoCrime
WHERE LOCATION_DESCRIPTION IS NULL;
"
# Thực thi truy vấn
result <- dbGetQuery(conn, query)
# In kết quả
print(result)
```

	total_cases <chr>
1	0

1 row

Problem 7 **How does the trend of total crime incidents change over the years? Is there an increase or decrease?**

```{r}
query <- "
SELECT YEAR(DATE), COUNT(*) AS total_cases
FROM ChicagoCrime
GROUP BY YEAR(DATE)
ORDER BY year
"
# Thực thi truy vấn
result <- dbGetQuery(conn, query)
# In kết quả
print(result)
```

	YEAR(DATE) <dbl>	total_cases <chr>
1	2001	36
2	2002	37
3	2003	42
4	2004	33
5	2005	44
6	2006	30
7	2007	36
8	2008	21
9	2009	43
10	2010	25

1-10 of 18 rows

Previous 1 2 Next

Problem 8 **Which types of crimes are increasing or decreasing over time? Is there a specific type that is particularly prevalent in recent years?**

```{r}
query <- "
SELECT YEAR(DATE), PRIMARY_TYPE , COUNT(*) AS total_crime
FROM ChicagoCrime
GROUP BY YEAR(DATE), PRIMARY_TYPE
ORDER BY YEAR, total_crime DESC
"
# Thực thi truy vấn
result <- dbGetQuery(conn, query)
# In kết quả
print(result)
```

	YEAR(DATE) <dbl>	PRIMARY_TYPE <chr>	total_crime <chr>
1	2001	BATTERY	10
2	2001	THEFT	5
3	2001	DECEPTIVE PRACTICE	4
4	2001	PROSTITUTION	3
5	2001	NARCOTICS	3
6	2001	CRIMINAL DAMAGE	2
7	2001	CRIMINAL TRESPASS	2
8	2001	MOTOR VEHICLE THEFT	2
9	2001	BURGLARY	2
10	2001	CRIM SEXUAL ASSAULT	1

1-10 of 206 rows

Previous 1 2 3 4 5 6 … 21 Next

Problem 9 **Which areas have a higher crime trend compared to others? Does this trend change over time?**

```{r}
query <- "
SELECT BLOCK, YEAR(DATE) , COUNT(*) AS total_crime
FROM ChicagoCrime
GROUP BY BLOCK, YEAR(DATE)
ORDER BY total_crime, YEAR(DATE) ASC
"
# Thực thi truy vấn
result <- dbGetQuery(conn, query)
# In kết quả
print(result)
```

	BLOCK <chr>	YEAR(DATE) <dbl>	total_crime <chr>
1	0000X S STATE ST	2001	1
2	063XX N NAGLE AV	2001	1
3	011XX W WILSON AV	2001	1
4	010XX N ORLEANS ST	2001	1
5	024XX W HADDON AV	2001	1
6	069XX S EAST END AV	2001	1
7	004XX W WRIGHTWOOD AV	2001	1
8	035XX W DICKENS AV	2001	1
9	055XX S NOTTINGHAM AV	2001	1
10	016XX E 95 ST	2001	1

1-10 of 531 rows

Previous 1 2 3 4 5 6 … 54 Next

Problem 10 **Is there a particular time of day (morning, afternoon, evening) when crimes are more likely to occur?**

```{r}
query <- "SELECT
  CASE
    WHEN HOUR(DATE) < 12 THEN 'Morning'
    WHEN HOUR(DATE) < 18 THEN 'Afternoon'
    ELSE 'Evening'
  END AS time_of_day,
  COUNT(*) AS total_cases
FROM ChicagoCrime
GROUP BY time_of_day
ORDER BY FIELD(time_of_day, 'Morning', 'Afternoon', 'Evening')
"
# Thực thi truy vấn
result <- dbGetQuery(conn, query)
# In kết quả
print(result)
```

	time_of_day <chr>	total_cases <chr>
1	Morning	533

1 row

Problem 11 **Is the crime trend affected by the season of the year? Which season typically has the highest number of incidents?**

```{r}
query <- "
SELECT
  CASE
    WHEN MONTH(DATE) IN (12, 1, 2) THEN 'Winter'
    WHEN MONTH(DATE) IN (3, 4, 5) THEN 'Spring'
    WHEN MONTH(DATE) IN (6, 7, 8) THEN 'Summer'
    WHEN MONTH(DATE) IN (9, 10, 11) THEN 'Fall'
  END AS Season,
  COUNT(*) AS total_cases
FROM ChicagoCrime
WHERE DATE IS NOT NULL
GROUP BY Season
ORDER BY total_cases DESC
"
# Thực thi truy vấn
result <- dbGetQuery(conn, query)
# In kết quả
print(result)
```

	Season <chr>	total_cases <chr>
1	Spring	148
2	Summer	140
3	Fall	137
4	Winter	108

4 rows

Problem 12 **Which type of crime is the most prevalent in the entire dataset? What is the frequency of that type over the years?**

```{r}
query <- "
SELECT
  PRIMARY_TYPE,
  YEAR(DATE) ,
  COUNT(*) AS FREQUENCY
FROM ChicagoCrime
GROUP BY PRIMARY_TYPE, YEAR(DATE)
ORDER BY YEAR(DATE)  ,FREQUENCY DESC
"
# Thực thi truy vấn
result <- dbGetQuery(conn, query)
# In kết quả
print(result)
```

	PRIMARY_TYPE <chr>	YEAR(DATE) <dbl>	FREQUENCY <chr>
1	BATTERY	2001	10
2	THEFT	2001	5
3	DECEPTIVE PRACTICE	2001	4
4	PROSTITUTION	2001	3
5	NARCOTICS	2001	3
6	CRIMINAL DAMAGE	2001	2
7	CRIMINAL TRESPASS	2001	2
8	BURGLARY	2001	2
9	MOTOR VEHICLE THEFT	2001	2
10	PUBLIC PEACE VIOLATION	2001	1

1-10 of 206 rows

Previous 1 2 3 4 5 6 … 21 Next

Problem 13 **Which locations have the highest frequency of crimes? Are there any notable patterns or hotspots?**

```{r}
query <- "
SELECT
  LOCATION_DESCRIPTION,
  COUNT(*) AS total_cases
FROM ChicagoCrime
GROUP BY LOCATION_DESCRIPTION
ORDER BY total_cases DESC
LIMIT 10
"
# Thực thi truy vấn
result <- dbGetQuery(conn, query)
# In kết quả
print(result)
```

	LOCATION_DESCRIPTION <chr>	total_cases <chr>
1	STREET	136
2	RESIDENCE	84
3	SIDEWALK	64
4	APARTMENT	59
5	OTHER	18
6	PARKING LOT/GARAGE(NON.RESID.)	16
7	RESTAURANT	11
8	ALLEY	11
9	RESIDENCE PORCH/HALLWAY	10
10	RESIDENCE-GARAGE	10

1-10 of 10 rows

Problem 14 **Is there a correlation between total incidents and arrest frequency? What percentage of incidents result in arrests?**

```{r}
query <- "
SELECT
  COUNT(*) AS total_incidents,
  SUM(CASE WHEN ARREST = TRUE THEN 1 ELSE 0 END) AS total_arrests,
  SUM(CASE WHEN ARREST = TRUE THEN 1 ELSE 0 END)/COUNT(*) as total_incidents_percent
FROM ChicagoCrime
"
# Thực thi truy vấn
result <- dbGetQuery(conn, query)
# In kết quả
print(result)
```

	total_incidents <chr>	total_arrests <chr>	total_incidents_percent <chr>
1	533	0	0.0000

1 row

Problem 15 **Which community areas have the highest frequency of crimes? Does this frequency change over time?**

```{r}
query <- "
SELECT
    COMMUNITY_AREA_NUMBER,
    COUNT(*) AS TotalCases
FROM
    ChicagoCrime
GROUP BY
    COMMUNITY_AREA_NUMBER
ORDER BY
    TotalCases DESC;
"
# Thực thi truy vấn
result <- dbGetQuery(conn, query)
# In kết quả
print(result)
```

	COMMUNITY_AREA_NUMBER <dbl>	TotalCases <chr>
1	NA	43
2	25	43
3	23	22
4	68	21
5	28	16
6	29	16
7	8	15
8	71	14
9	24	13
10	67	12

1-10 of 71 rows

Previous 1 2 3 4 5 6 … 8 Next

Problem 16 **Which location descriptions have the highest frequency of crimes? Is there a difference between various location descriptions?**

```{r}
query <- "
SELECT
    LOCATION_DESCRIPTION,
    COUNT(*) AS TotalCases
FROM
    ChicagoCrime
GROUP BY
    LOCATION_DESCRIPTION
ORDER BY
    TotalCases DESC;
"
# Thực thi truy vấn
result <- dbGetQuery(conn, query)
# In kết quả
print(result)
```

	LOCATION_DESCRIPTION <chr>	TotalCases <chr>
1	STREET	136
2	RESIDENCE	84
3	SIDEWALK	64
4	APARTMENT	59
5	OTHER	18
6	PARKING LOT/GARAGE(NON.RESID.)	16
7	RESTAURANT	11
8	ALLEY	11
9	RESIDENCE PORCH/HALLWAY	10
10	RESIDENCE-GARAGE	10

1-10 of 48 rows

Previous 1 2 3 4 5 Next

Problem 17 **Which types of crimes are more prevalent in specific locations? Does this trend change over time?**

```{r}
query <- "
SELECT
    LOCATION_DESCRIPTION,
    PRIMARY_TYPE,
    COUNT(*) AS TotalCases
FROM
    ChicagoCrime
GROUP BY
    LOCATION_DESCRIPTION, PRIMARY_TYPE
ORDER BY
    LOCATION_DESCRIPTION, TotalCases DESC;
"
# Thực thi truy vấn
result <- dbGetQuery(conn, query)
# In kết quả
print(result)
```

	LOCATION_DESCRIPTION <chr>	PRIMARY_TYPE <chr>	TotalCases <chr>
1	AIRPORT/AIRCRAFT	THEFT	1
2	AIRPORT/AIRCRAFT	NARCOTICS	1
3	AIRPORT/AIRCRAFT	DECEPTIVE PRACTICE	1
4	ALLEY	NARCOTICS	3
5	ALLEY	CRIM SEXUAL ASSAULT	2
6	ALLEY	MOTOR VEHICLE THEFT	1
7	ALLEY	ROBBERY	1
8	ALLEY	LIQUOR LAW VIOLATION	1
9	ALLEY	CRIMINAL TRESPASS	1
10	ALLEY	ASSAULT	1

1-10 of 177 rows