

R Notebook

Code

Hide

```
#install.packages("odbc")
#install.packages("readr")
#install.packages("R00BC")
#install.packages('DBI')
```

1. Install packages and set up libraries

Hide

```
library(DBI) # Để sử dụng dbWriteTable
library(odbc) # Nếu bạn kết nối bằng ODBC
library(readr) # Để đọc file CSV
```

2. Connect to the data server

Hide

```
server <- "localhost,1433" # Địa chỉ máy chủ SQL Server
database <- "DSR_demo" # Tên cơ sở dữ liệu
user <- "sa" # Tên người dùng
password <- "kiet123654" # Mật khẩu

# Tạo chuỗi kết nối
connection_string <- paste0("Driver={SQL Server};",
                             "Server=", server, ";",
                             "Database=", database, ";",
                             "Uid=", user, ";",
                             "Pwd=", password, ";",
                             "charset=utf-8") # Tạo kết nối hỗ trợ Unicode
```

Hide

```
con <- dbConnect(odbc::odbc(), .connection_string = connection_string)

# Kiểm tra kết nối
if (is.null(con)) {
  cat("Kết nối không thành công.\n")
} else {
  cat("Kết nối thành công đến cơ sở dữ liệu ", database, ".\n")
}
```

Kết nối thành công đến cơ sở dữ liệu DSR_demo .

3. Import data into the database

Hide

```
#chicago_data <- read.csv("ChicagoPublicSchool.csv", stringsAsFactors = FALSE)

dbWriteTable(con,
             name="ChIPS",
             value=chicago_data,
             overwrite=TRUE)
```

Hide

```
# Kiểm tra số lượng bản ghi đã được nhập
row_count <- dbGetQuery(con, "SELECT COUNT(*) AS count
                             FROM ChIPS")
cat("Số lượng bản ghi đã được nhập:", row_count$count, "\n")
```

Số lượng bản ghi đã được nhập: 566

Hide

```
# In ra thông báo khi quá trình nhập dữ liệu hoàn thành
cat("Quá trình nhập dữ liệu đã hoàn thành thành công!\n")
```

Quá trình nhập dữ liệu đã hoàn thành thành công!

4. Summarize the data

Hide

```
chicago_data <- dbReadTable(con, "ChIPS")
summary(chicago_data)
```

School_ID	NAME_OF_SCHOOL	Elementary_Middle_or_High_School	Street_Address	
Min. :400018	Length:566	Length:566	Length:566	
1st Qu.:609873	Class :character	Class :character	Class :character	
Median :610053	Mode :character	Mode :character	Mode :character	
Mean :609682				
3rd Qu.:610213				
Max. :610544				
City	State	ZIP_Code	Phone_Number	Link
Length:566	Length:566	Min. :60605	Length:566	Length:566
Class :character	Class :character	1st Qu.:60618	Class :character	Class :character
Mode :character	Mode :character	Median :60625	Mode :character	Mode :character
		Mean :60630		
		3rd Qu.:60639		
		Max. :60827		
Network_Manager	Collaborative_Name	Adequate_Yearly_Progress_Made	Track_Schedule	
Length:566	Length:566	Mode :logical	Length:566	
Class :character	Class :character	FALSE:476	Class :character	
Mode :character	Mode :character	TRUE :72	Mode :character	
		NA's :18		
CPS_Performance_Policy_Status	CPS_Performance_Policy_Level	HEALTHY_SCHOOL_CERTIFIED		
Length:566	Length:566	Mode :logical		
Class :character	Class :character	FALSE:550		
Mode :character	Mode :character	TRUE :16		
Safety_Icon	SAFETY_SCORE	Family_Involvement_Icon	Family_Involvement_Score	
Length:566	Min. :1.0	Length:566	Min. :6.0	
Class :character	1st Qu.:35.0	Class :character	1st Qu.:37.0	
Mode :character	Median :48.0	Mode :character	Median :49.0	
	Mean :49.5		Mean :50.6	
	3rd Qu.:61.0		3rd Qu.:61.0	
	Max. :99.0		Max. :99.0	
	NA's :53		NA's :269	
Environment_Icon	Environment_Score	Instruction_Icon	Instruction_Score	Leaders_Icon
Length:566	Min. :1.00	Length:566	Min. :1.00	Length:566
Class :character	1st Qu.:37.00	Class :character	1st Qu.:37.00	Class :character
Mode :character	Median :47.00	Mode :character	Median :47.00	Mode :character
	Mean :47.77		Mean :48.29	
	3rd Qu.:58.00		3rd Qu.:59.00	
	Max. :99.00		Max. :99.00	
	NA's :53		NA's :53	

Leaders_Score	Teachers_Icon	Teachers_Score	Parent_Engagement_Icon	
Min. : 4.00	Length:566	Min. : 6.00	Length:566	
1st Qu.:35.50	Class :character	1st Qu.:36.00	Class :character	
Median :49.00	Mode :character	Median :48.00	Mode :character	
Mean :49.89		Mean :49.05		
3rd Qu.:62.00		3rd Qu.:59.00		
Max. :99.00		Max. :99.00		
NA's :271		NA's :271		
Parent_Engagement_Score	Parent_Environment_Icon	Parent_Environment_Score		
Min. :37.00	Length:566	Min. :31.00		
1st Qu.:47.00	Class :character	1st Qu.:47.00		
Median :50.00	Mode :character	Median :50.00		
Mean :50.23		Mean :50.19		
3rd Qu.:53.00		3rd Qu.:53.00		
Max. :69.00		Max. :70.00		
NA's :134		NA's :134		
AVERAGE_STUDENT_ATTENDANCE	Rate_of_Misconducts_per_100_students	Average_Teacher_Attendance		
Length:566	Min. : 0.000	Length:566		
Class :character	1st Qu.: 4.525	Class :character		
Mode :character	Median : 12.250	Mode :character		
	Mean : 21.086			
	3rd Qu.: 26.975			
	Max. :251.600			
Individualized_Education_Program_Compliance_Rate	Pk_2_Literacy	Pk_2_Math		
Length:566	Min. : 16.40	Min. : 6.90		
Class :character	1st Qu.: 53.80	1st Qu.:34.70		
Mode :character	Median : 64.70	Median :44.40		
	Mean : 63.80	Mean :44.71		
	3rd Qu.: 73.85	3rd Qu.:55.10		
	Max. :100.00	Max. :97.20		
	NA's :163	NA's :225		
Gr3_5_Grade_Level_Math	Gr3_5_Grade_Level_Read	Gr3_5_Keep_Pace_Read	Gr3_5_Keep_Pace_Math	
Min. : 0.00	Min. : 0.00	Min. :27.30	Min. :12.80	
1st Qu.: 23.68	1st Qu.: 18.98	1st Qu.:44.88	1st Qu.:40.20	
Median : 35.55	Median : 27.90	Median :51.65	Median :56.30	
Mean : 37.95	Mean : 33.36	Mean :51.75	Mean :56.13	
3rd Qu.: 47.75	3rd Qu.: 42.62	3rd Qu.:57.90	3rd Qu.:63.90	
Max. :100.00	Max. :100.00	Max. :82.50	Max. :83.50	
NA's :114	NA's :114	NA's :114	NA's :114	
Gr6_8_Grade_Level_Math	Gr6_8_Grade_Level_Read	Gr6_8_Keep_Pace_Math	Gr6_8_Keep_Pace_Read	
Min. : 0.00	Min. : 2.90	Min. :22.40	Min. :20.20	
1st Qu.: 30.10	1st Qu.: 25.40	1st Qu.:49.20	1st Qu.:48.20	
Median : 41.35	Median : 34.20	Median :56.90	Median :54.30	
Mean : 42.76	Mean : 38.72	Mean :56.79	Mean :54.11	
3rd Qu.: 54.05	3rd Qu.: 49.10	3rd Qu.:64.42	3rd Qu.:60.70	
Max. :100.00	Max. :100.00	Max. :93.60	Max. :85.50	
NA's :122	NA's :121	NA's :122	NA's :121	
Gr_8_Explore_Math	Gr_8_Explore_Read	ISAT_Exceeding_Math	ISAT_Exceeding_Reading	
Min. : 0.00	Min. : 0.00	Min. : 0.000	Min. : 0.00	
1st Qu.: 6.40	1st Qu.: 15.10	1st Qu.: 9.175	1st Qu.: 6.10	
Median : 12.30	Median : 25.20	Median : 16.000	Median : 10.95	
Mean : 16.88	Mean : 28.31	Mean : 20.356	Mean : 15.67	
3rd Qu.: 21.60	3rd Qu.: 36.80	3rd Qu.: 24.900	3rd Qu.: 19.52	
Max. :100.00	Max. :100.00	Max. :100.000	Max. :100.00	
NA's :141	NA's :141	NA's :90	NA's :90	
ISAT_Value_Add_Math	ISAT_Value_Add_Read	ISAT_Value_Add_Color_Math	ISAT_Value_Add_Color_Read	
Min. : -3.50000	Min. : -5.00000	Length:566	Length:566	
1st Qu.: -0.60000	1st Qu.: -0.60000	Class :character	Class :character	
Median : 0.00000	Median : 0.00000	Mode :character	Mode :character	
Mean : 0.06859	Mean : 0.02265			
3rd Qu.: 0.70000	3rd Qu.: 0.62500			
Max. : 3.60000	Max. : 4.90000			
NA's :98	NA's :98			
Students_Taking_Algebra	Students_Passing_Algebra	X_9th_Grade_EXPLORE_2009		
Length:566	Length:566	Length:566		
Class :character	Class :character	Class :character		
Mode :character	Mode :character	Mode :character		
X_9th_Grade_EXPLORE_2010	X_10th_Grade_PLAN_2009	X_10th_Grade_PLAN_2010		
Length:566	Length:566	Length:566		
Class :character	Class :character	Class :character		
Mode :character	Mode :character	Mode :character		
Net_Change_EXPLORE_and_PLAN	X_11th_Grade_Average_ACT_2011	Net_Change_PLAN_and_ACT		
Length:566	Length:566	Length:566		
Class :character	Class :character	Class :character		
Mode :character	Mode :character	Mode :character		
College_Eligibility	Graduation_Rate	College_Enrollment_Rate	COLLEGE_ENROLLMENT	
Length:566	Length:566	Length:566	Min. : 21.0	
Class :character	Class :character	Class :character	1st Qu.: 342.2	
Mode :character	Mode :character	Mode :character	Median : 504.0	
			Mean : 626.1	
			3rd Qu.: 790.5	
			Max. :4368.0	
General_Services_Route	Freshman_on_Track_Rate	X_COORDINATE	Y_COORDINATE	
Min. :29.00	Length:566	Min. :1118114	Min. :1817242	
1st Qu.:34.00	Class :character	1st Qu.:1152166	1st Qu.:1860248	
Median :39.00	Mode :character	Median :1163553	Median :1866889	
Mean :39.08		Mean :1163016	Mean :1885663	
3rd Qu.:45.00		3rd Qu.:1174256	3rd Qu.:1911652	
Max. :49.00		Max. :1202811	Max. :1950960	
Latitude	Longitude	COMMUNITY_AREA_NUMBER	COMMUNITY_AREA_NAME	Ward
Min. :41.65	Min. : -87.84	Min. : 1.00	Length:566	Min. : 1.00
1st Qu.:41.77	1st Qu.: -87.72	1st Qu.:23.00	Class :character	1st Qu.:10.00
Median :41.85	Median : -87.68	Median :31.50	Mode :character	Median :21.00
Mean :41.84	Mean : -87.68	Mean :37.71		Mean :22.06
3rd Qu.:41.91	3rd Qu.: -87.64	3rd Qu.:58.00		3rd Qu.:32.75
Max. :42.02	Max. : -87.53	Max. :77.00		Max. :50.00
Police_District	Location			
Min. : 1.00	Length:566			
1st Qu.: 7.00	Class :character			
Median :10.00	Mode :character			
Mean :11.71				
3rd Qu.:17.00				
Max. :25.00				

5. Retrieve table structure information

columns_query <- dbGetQuery(con, " SELECT School_ID FROM ChIPS WHERE School_ID = 610281") print(columns_query)	Hide
	School_ID
	<in>
	610281

1 row

6. Rename columns in the `chicago_public_schools` table

Hide

```
dbExecute(con, "
    EXEC sp_rename 'ChIPS.[Elementary_Middle_or_High_School]', 'School_Level', 'COLUMN';
")
```

[1] 0

7. Count the number of elementary schools

Hide

```
query_elementary_count <- dbGetQuery(con, "
    SELECT COUNT(School_Level) AS count
    FROM ChIPS
    WHERE School_Level = 'ES';
")
print(query_elementary_count)
```

	count
	<int>
	462

1 row

8. Find the highest safety score

Hide

```
query_highest_safety_score <- dbGetQuery(con, "
    SELECT MAX(SAFETY_SCORE) AS highest_safety_score
    FROM ChIPS
")
print(query_highest_safety_score)
```

	highest_safety_score
	<int>
	99

1 row

9. Write R code to clean the column names of a data frame called `chicago_schools`.

Hide

```
clean_column_names <- function(names) {
  names <- gsub("_+", "_", names)

  names <- gsub("[_\\.]+$", "", names)

  names <- gsub("\\.", "_", names)

  return(names)
}
```

10. Check for missing data

Hide

```
missing_data <- colSums(is.na(chicago_data))
print(missing_data)
```

```

      School_ID
      0
NAME_OF_SCHOOL
      0
Elementary_Middle_or_High_School
      0
      Street_Address
      0
      City
      0
      State
      0
      ZIP_Code
      0
      Phone_Number
      0
      Link
      1
      Network_Manager
      0
      Collaborative_Name
      0
      Adequate_Yearly_Progress_Made
      18
      Track_Schedule
      0
      CPS_Performance_Policy_Status
      0
      CPS_Performance_Policy_Level
      0
      HEALTHY_SCHOOL_CERTIFIED
      0
      Safety_Icon
      0
      SAFETY_SCORE
      53
      Family_Involvement_Icon
      0
      Family_Involvement_Score
      269
      Environment_Icon
      0
      Environment_Score
      53
      Instruction_Icon
      0
      Instruction_Score
      53
      Leaders_Icon
      0
      Leaders_Score
      271
      Teachers_Icon
      0
      Teachers_Score
      271
      Parent_Engagement_Icon
      0
      Parent_Engagement_Score
      134
      Parent_Environment_Icon
      0
      Parent_Environment_Score
```

```

134
AVERAGE_STUDENT_ATTENDANCE
1
Rate_of_Misconducts_per_100_students
0
Average_Teacher_Attendance
0
Individualized_Education_Program_Compliance_Rate
0
Pk_2_Literacy
163
Pk_2_Math
225
Gr3_5_Grade_Level_Math
114
Gr3_5_Grade_Level_Read
114
Gr3_5_Keep_Pace_Read
114
Gr3_5_Keep_Pace_Math
114
Gr6_8_Grade_Level_Math
122
Gr6_8_Grade_Level_Read
121
Gr6_8_Keep_Pace_Math
122
Gr6_8_Keep_Pace_Read
121
Gr_8_Explore_Math
141
Gr_8_Explore_Read
141
ISAT_Exceeding_Math
90
ISAT_Exceeding_Reading
90
ISAT_Value_Add_Math
98
ISAT_Value_Add_Read
98
ISAT_Value_Add_Color_Math
0
ISAT_Value_Add_Color_Read
0
Students_Taking_Algebra
0
Students_Passing_Algebra
0
X_9th_Grade_EXPLORE_2009
0
X_9th_Grade_EXPLORE_2010
0
X_10th_Grade_PLAN_2009
0
X_10th_Grade_PLAN_2010
0
Net_Change_EXPLORE_and_PLAN
0
X_11th_Grade_Average_ACT_2011
0
Net_Change_PLAN_and_ACT
0
College_Eligibility
0
Graduation_Rate
0
College_Enrollment_Rate
0
COLLEGE_ENROLLMENT
0
General_Services_Route
0
Freshman_on_Track_Rate
0
X_COORDINATE
0
Y_COORDINATE
0
Latitude
0
Longitude
0
COMMUNITY_AREA_NUMBER
0
COMMUNITY_AREA_NAME
0
Ward
0
Police_District
0
Location
0

```