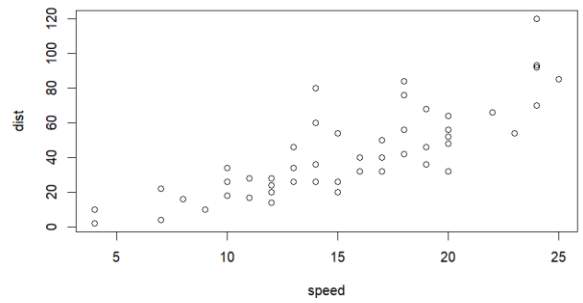
SE183091_Nguyễn Thanh Hòa _Test_2

```
plot(cars)
```




```{r}
install.packages("RODBC")
install.packages("odbc")
install.packages('readr')
install.packages("DBI")
```

```{r}
library(RODBC)
library(odbc)
library(readr)
library(DBI)
```

```{r}
# Thay đổi các thông số kết nối tùy theo cài đặt của bạn
server <- "localhost, 1433"  # Địa chỉ máy chủ SQL Server
database <- "ChicagoPublicSchools_test1"       # Tên cơ sở dữ liệu
user <- "sa"                  # Tên người dùng
password <- "123456789"    # Mật khẩu

connection_string <- paste0("Driver={SQL Server};",
                            "Server=", server, ";",
                            "Database=", database, ";",
                            "Uid=", user, ";",
                            "Pwd=", password, ";",
                            "charset=UTF-8")

# Kết nối đến cơ sở dữ liệu
con <- dbConnect(odbc::odbc(), .connection_string = connection_string)

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

```
Kết nối thành công đến cơ sở dữ liệu  ChicagoPublicSchools_test1 .
```

```{r}
# Đọc file CSV
chicago_schools <- read.csv("ChicagoPublicSchools.csv")


dbWriteTable(con, "chicago_public_schools", chicago_schools, overwrite =
TRUE)

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

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

```
Số lượng bản ghi đã được nhập: 566
Quá trình nhập dữ liệu đã hoàn thành thành công!
```

1. List the top 5 schools with the highest graduation rates

```r
# Chuyển đổi cột Graduation_Rate sang kiểu số
chicago_schools_filtered$Graduation_Rate <- as.numeric(chicago_schools_filtered$Graduation_Rate)
```

```
Warning: NAs introduced by coercion
```

Hide

```r
# Kiểm tra nếu có giá trị không thể chuyển đổi thành số
if (any(is.na(chicago_schools_filtered$Graduation_Rate))) {
  warning("Có giá trị NA trong cột Graduation_Rate sau khi chuyển đổi sang kiểu số.")
}
```

```
Warning: Có giá trị NA trong cột Graduation_Rate sau khi chuyển đổi sang kiểu số.
```

```r
# Sắp xếp các trường theo tỷ lệ tốt nghiệp giảm dần và lấy top 5
top_graduation_schools <- chicago_schools_filtered[order(-chicago_schools_filtered$Graduation_Rate), ]
top_5_schools <- head(top_graduation_schools, 5)

# In ra top 5 trường
print(top_5_schools)
```

	School_ID	NAME_OF_SCHOOL	Elementary..Middle..or.High.School
	<int>	<chr>	<chr>
409	609749	Northside College Preparatory High School	HS
519	609680	Walter Payton College Preparatory High School	HS
532	609755	Whitney M Young Magnet High School	HS
7	609720	Albert G Lane Technical High School	HS
554	609678	William Jones College Preparatory High School	HS

5 rows | 1-4 of 79 columns

Hide

```
NA
```

2. Calculate the average safety score for each school type

```r
library(dplyr)
```

```
colnames(chicago_schools)
```

```
 [1] "School_ID"
 [2] "NAME_OF_SCHOOL"
 [3] "Elementary..Middle..or.High.School"
 [4] "Street_Address"
 [5] "City"
 [6] "State"
 [7] "ZIP_Code"
 [8] "Phone_Number"
 [9] "Link"
[10] "Network_Manager"
[11] "Collaborative_Name"
[12] "Adequate_Yearly_Progress_Made_"
[13] "Track_Schedule"
[14] "CPS_Performance_Policy_Status"
[15] "CPS_Performance_Policy_Level"
[16] "HEALTHY_SCHOOL_CERTIFIED"
[17] "Safety_Icon"
[18] "SAFETY_SCORE"
[19] "Family_Involvement_Icon"
[20] "Family_Involvement_Score"
[21] "Environment_Icon"
[22] "Environment_Score"
[23] "Instruction_Icon"
[24] "Instruction_Score"
[25] "Leaders_Icon"
[26] "Leaders_Score"
[27] "Teachers_Icon"
[28] "Teachers_Score"
[29] "Parent_Engagement_Icon"
[30] "Parent_Engagement_Score"
[31] "Parent_Environment_Icon"
```

```
[32] "Parent_Environment_Score"
[33] "AVERAGE_STUDENT_ATTENDANCE"
[34] "Rate_of_Misconducts__per_100_students_"
[35] "Average_Teacher_Attendance"
[36] "Individualized_Education_Program_Compliance_Rate"
[37] "Pk_2_Literacy__"
[38] "Pk_2_Math__"
[39] "Gr3_5_Grade_Level_Math__"
[40] "Gr3_5_Grade_Level_Read__"
[41] "Gr3_5_Keep_Pace_Read__"
[42] "Gr3_5_Keep_Pace_Math__"
[43] "Gr6_8_Grade_Level_Math__"
[44] "Gr6_8_Grade_Level_Read__"
[45] "Gr6_8_Keep_Pace_Math_"
[46] "Gr6_8_Keep_Pace_Read__"
[47] "Gr_8_Explore_Math__"
[48] "Gr_8_Explore_Read__"
[49] "ISAT_Exceeding_Math__"
[50] "ISAT_Exceeding_Reading__"
[51] "ISAT_Value_Add_Math"
[52] "ISAT_Value_Add_Read"
[53] "ISAT_Value_Add_Color_Math"
[54] "ISAT_Value_Add_Color_Read"
[55] "Students_Taking__Algebra__"
[56] "Students_Passing__Algebra__"
[57] "X9th.Grade.EXPLORE..2009."
[58] "X9th.Grade.EXPLORE..2010."
[59] "X10th.Grade.PLAN..2009."
[60] "X10th.Grade.PLAN..2010."
[61] "Net_Change_EXPLORE_and_PLAN"
[62] "X11th.Grade.Average.ACT..2011."
[63] "Net_Change_PLAN_and_ACT"
[64] "College_Eligibility__"
```

```
[65] "Graduation_Rate__"
[66] "College_Enrollment_Rate__"
[67] "COLLEGE_ENROLLMENT"
[68] "General_Services_Route"
[69] "Freshman_on_Track_Rate__"
[70] "X_COORDINATE"
[71] "Y_COORDINATE"
[72] "Latitude"
[73] "Longitude"
[74] "COMMUNITY_AREA_NUMBER"
[75] "COMMUNITY_AREA_NAME"
[76] "Ward"
[77] "Police_District"
[78] "Location"
```

```{r}
colnames(chicago_schools)

average_safety_scores <- chicago_schools %>%
  group_by(Elementary..Middle..or.High.School) %>%
  summarise(avg_safety_score = mean(SAFETY_SCORE, na.rm = TRUE)) %>%
  arrange(desc(avg_safety_score))
# In ra kết quả
print(average_safety_scores)

```

A tibble: 3 × 2

Elementary..Middle..or.High.School <chr>	avg_safety_score <dbl>
HS	49.62353
ES	49.52038
MS	48.00000

3 rows

3. Count the number of "Healthy School" certified schools

```{r}
healthy_school_count <- chicago_schools %>%
  filter(HEALTHY_SCHOOL_CERTIFIED == "Yes") %>%
  summarise(count = n())

# In ra kết quả
print(healthy_school_count)
```

Description: df [1 × 1]

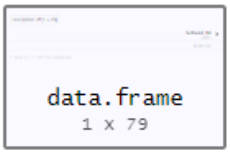
count <int>
16

1 row

4. Find the school with the highest percentage of students taking Algebra

```{r}
highest_algebra_participation <- chicago_schools %>%
  filter(!is.na(Students_Taking__Algebra__)) %>%
  mutate(Participation_Rate = as.numeric(Students_Taking__Algebra__)) %>%
  arrange(desc(Participation_Rate)) %>%
  slice(1)

# In ra kết quả
print(highest_algebra_participation)
```



Description: df [1 × 79]

School_ID <int>	NAME_OF_SCHOOL <chr>
610122	Helen Peirce International Studies Elementary School

1 row | 1-2 of 79 columns

5. Calculate the average ACT score for high schools

```{r}
average_act_score <- chicago_schools %>%
  filter(Elementary..Middle..or.High.School == "HS") %>%
  mutate(ACT_Score = as.numeric(ifelse(X11th.Grade.Average.ACT..2011. ==
"NDA", NA, X11th.Grade.Average.ACT..2011.))) %>%
  summarise(avg_ACT = mean(ACT_Score, na.rm = TRUE))

# In ra kết quả
print(average_act_score)

```

Description: df [1 × 1]

avg_ACT <dbl>
16.8012

1 row

6. Count the number of schools in each community area

```{r}
schools_per_community <- chicago_schools %>%
  group_by(COMMUNITY_AREA_NAME) %>%
  summarise(Number_of_Schools = n())

# In ra kết quả
print(schools_per_community)
```

A tibble: 77 × 2

COMMUNITY_AREA_NAME <chr>	Number_of_Schools <int>
ALBANY PARK	8
ARCHER HEIGHTS	2
ARMOUR SQUARE	3
ASHBURN	8
AUBURN GRESHAM	10
AUSTIN	23
AVALON PARK	3
AVONDALE	4
BELMONT CRAGIN	12
BEVERLY	4

1-10 of 77 rows Previous 1 2 3 4 5 6 ... 8 Next

7. Identify the school with the highest college enrollment rate

```{r}
highest_college_enrollment <- chicago_schools %>%
  filter(!is.na(COLLEGE_ENROLLMENT)) %>%
  arrange(desc(COLLEGE_ENROLLMENT)) %>%
  slice(1)

# In ra kết quả
print(highest_college_enrollment)
```

Description: df [1 × 78]

School_ID <int>	NAME_OF_SCHOOL <chr>
609720	Albert G Lane Technical High School

1 row | 1-2 of 78 columns

#8. Calculate the average environment score for each Network Manager

```r
average_environment_score <- chicago_schools %>%
  group_by(Network_Manager) %>%
  summarise(avg_environment_score = mean(Environment_Score, na.rm = TRUE)) %>%
  arrange(desc(avg_environment_score))

# In ra kết quả
print(average_environment_score)
```

A tibble: 20 × 2

Network_Manager <chr>	avg_environment_score <dbl>
Ravenswood-Ridge Elementary Network	55.89474
AUSL Schools	55.05882
O'Hare Elementary Network	52.16216
Austin-North Lawndale Elementary Network	51.89286
West Side High School Network	51.00000
North-Northwest Side High School Network	50.04348
Garfield-Humboldt Elementary Network	50.00000
Fulton Elementary Network	49.92593
Fullerton Elementary Network	49.26471
Skyway Elementary Network	47.96875

1-10 of 20 rows Previous 1 2 Next

9. Count schools achieving "Level 1" in CPS Performance Policy

```r
level_1_count <- chicago_schools %>%
  filter(CPS_Performance_Policy_Level == "Level 1") %>%
  summarise(Number_of_Level_1_Schools = n())

# In ra kết quả
print(level_1_count)
```

Description: df [1 × 1]

Number_of_Level_1_Schools <int>
139

1 row

10. Calculate total college enrollment by community area

```r
college_enrollment_by_community <- chicago_schools %>%
  group_by(COMMUNITY_AREA_NAME) %>%
  summarise(total_college_enrollment = sum(COLLEGE_ENROLLMENT, na.rm = TRUE)) %>%
  arrange(desc(total_college_enrollment))

# In ra kết quả
print(college_enrollment_by_community)
```

A tibble: 77 × 2

COMMUNITY_AREA_NAME <chr>	total_college_enrollment <int>
SOUTH LAWNDALE	14793
BELMONT CRAGIN	14386
AUSTIN	10933
GAGE PARK	9915
BRIGHTON PARK	9647
WEST TOWN	9429
HUMBOLDT PARK	8620
WEST RIDGE	8197
NEAR WEST SIDE	7975
NEW CITY	7922

1-10 of 77 rows Previous 1 2 3 4 5 6 … 8 Next