

Northeastern University

Course: DA5020
Assignment: Comparison between dplyr and SQL SELECT
Total Points: 100
Date Due: Posted on Blackboard

Learning Objectives

In this assignment, you will learn how to:

- Extract aggregated statistics using dplyr
- Compare dplyr to SQL SELECT

Tasks

- 1) Revisit your Bird Strike database created for the last assignment. Connect to the database. (10 points)

For the following functions, use dplyr operations to retrieve the necessary data.

- 2) Write a dplyr statement that counts the number of incidents where the incident reported fog during the incident. (10 points)
- 3) Write a function **CountIncidents(aircraftType)** that accepts an aircraft type and returns the number of bird strike incidents for the aircraft type. (20 points)
- 4) Write a function **Incidents(Airline)** that accepts an airline and returns a dataframe that contains all incidents for that airline. Limit the columns to: **AirportName**, **AircraftModel**, and **FlightDate**. (20 points)
- 5) Write a function **CountIncidentsByAirline()** that creates a data frame where the first column is a name of an Aircraft and the second column is the total number of incidents the Airline had. (20 points)
- 6) Use **show_query()** to extract the SQL statement from one of your dplyr operations, then compare it with the corresponding SQL statement you wrote last week. State your preference for retrieving data from a large data set. (20 points).

Complete the tasks and submit the R script to blackboard.