Module18

Joe Vargovich

11/10/2020

Exercise 1 - SQLite database for nycflights13

```
#a.Start up a SQL-Lite database with the NYCFlights13 data pre-loaded
con <- nycflights13_sqlite( )
```

```
## Caching nycflights db at C:\Users\josep\AppData\Local\Temp\RtmpoBQM7F/nycflights13.sqlite
```

```
## Creating table: airlines
```

```
## Creating table: airports
```

```
## Creating table: flights
```

```
## Creating table: planes
```

```
## Creating table: weather
```

```
#b. Connect to the flights and airlines tables.
flights = tbl(con, "flights")
airlines = tbl(con, "airlines")
head(flights)
```

```
## # Source:   lazy query [?? x 19]
## # Database: sqlite 3.33.0
## #   [C:\Users\josep\AppData\Local\Temp\RtmpoBQM7F\nycflights13.sqlite]
##    year month   day dep_time sched_dep_time dep_delay arr_time sched_arr_time
##   <int> <int> <int>    <int>          <int>     <dbl>    <int>          <int>
## 1  2013     1     1      517            515         2      830            819
## 2  2013     1     1      533            529         4      850            830
## 3  2013     1     1      542            540         2      923            850
## 4  2013     1     1      544            545        -1     1004           1022
## 5  2013     1     1      554            600        -6      812            837
## 6  2013     1     1      554            558        -4      740            728
## # ... with 11 more variables: arr_delay <dbl>, carrier <chr>, flight <int>,
## #   tailnum <chr>, origin <chr>, dest <chr>, air_time <dbl>, distance <dbl>,
## #   hour <dbl>, minute <dbl>, time_hour <dbl>
```

```
head(airlines)
```

```
## # Source:   lazy query [?? x 2]
## # Database: sqlite 3.33.0
## #   [C:\Users\josep\AppData\Local\Temp\RtmpoBQM7F\nycflights13.sqlite]
##   carrier name
##   <chr>   <chr>
## 1 9E      Endeavor Air Inc.
## 2 AA      American Airlines Inc.
```

```
## 3 AS      Alaska Airlines Inc.
## 4 B6      JetBlue Airways
## 5 DL      Delta Air Lines Inc.
## 6 EV      ExpressJet Airlines Inc.
```

```
#c. Percentage of late flights by carrier.
#Join our dataframes by carrier
combinedDf = inner_join(flights, airlines) %>%
  mutate(
    delayed = ifelse(dep_delay >= 10, "Late", "OnTime")
  )
```

```
## Joining, by = "carrier"
```

```
#Count occurances of ontime and delayed flights and find the percent late.
lateDf = combinedDf %>%
  as.data.frame() %>% #Needed to convert to a dataframe first before pivoting.
  count(name, delayed) %>%
  filter(!is.na(delayed)) %>%
  pivot_wider(names_from=delayed, values_from=n) %>%
  mutate(
    PercentLate = round(((Late)/(Late + OnTime) * 100), 2)
  )
lateDf
```

```
## # A tibble: 16 x 4
##    name                          Late OnTime PercentLate
##    <chr>                        <int>  <int>       <dbl>
##  1 AirTran Airways Corporation  1049   2138        32.9
##  2 Alaska Airlines Inc.          122    590        17.1
##  3 American Airlines Inc.       6218  25875        19.4
##  4 Delta Air Lines Inc.         9737  38024        20.4
##  5 Endeavor Air Inc.            5212  12204        29.9
##  6 Envoy Air                    6254  18909        24.8
##  7 ExpressJet Airlines Inc.    17972  33384        35.0
##  8 Frontier Airlines Inc.        241    441        35.3
##  9 Hawaiian Airlines Inc.         29    313         8.48
## 10 JetBlue Airways             14966  39203        27.6
## 11 Mesa Airlines Inc.            182    363        33.4
## 12 SkyWest Airlines Inc.           7     22        24.1
## 13 Southwest Airlines Co.       4037   8046        33.4
## 14 United Air Lines Inc.       15503  42476        26.7
## 15 US Airways Inc.              3022  16851        15.2
## 16 Virgin America               1142   3989        22.3
```

```
#d. Close the connection
dbDisconnect(con)
```

Exericse 2 - Create Sqlite database manually

```
#Here we create a .db file to hold the database created. We also connect to it.
con <- DBI::dbConnect(RSQLite::SQLite(), dbname = "TestSQLiteFile.db")

# Create a table using the iris data
dbCreateTable(con, 'IRIS', iris)
```

```r
head(iris)
```

```
##   Sepal.Length Sepal.Width Petal.Length Petal.Width Species
## 1          5.1         3.5          1.4         0.2  setosa
## 2          4.9         3.0          1.4         0.2  setosa
## 3          4.7         3.2          1.3         0.2  setosa
## 4          4.6         3.1          1.5         0.2  setosa
## 5          5.0         3.6          1.4         0.2  setosa
## 6          5.4         3.9          1.7         0.4  setosa
```

```r
dbListTables(con)
```

```
## [1] "IRIS"
```

```r
dbDisconnect(con)
```