Lesson 5: Chapter 10 and 11

Hao Wang

3/1/2022

0. Load libraries

```
knitr::opts_chunk$set(echo = TRUE)

library(tidyverse)
library(rlang)

# Load NYC flight dataset
library(nycflights13)
```

1. Data wrangling

Import -> Tidy -> Transform

2. Chapter 10. Tibbles: A new way of data frame in R

How to know if an object is a data frame or a tibble?

```
str(iris)
```

```
## 'data.frame':    150 obs. of  5 variables:
##  $ Sepal.Length: num  5.1 4.9 4.7 4.6 5 5.4 4.6 5 4.4 4.9 ...
##  $ Sepal.Width : num  3.5 3 3.2 3.1 3.6 3.9 3.4 3.4 2.9 3.1 ...
##  $ Petal.Length: num  1.4 1.4 1.3 1.5 1.4 1.7 1.4 1.5 1.4 1.5 ...
##  $ Petal.Width : num  0.2 0.2 0.2 0.2 0.2 0.4 0.3 0.2 0.2 0.1 ...
##  $ Species     : Factor w/ 3 levels "setosa","versicolor",..: 1 1 1 1 1 1 1 1 1 1 ...
```

```
class(iris)
```

```
## [1] "data.frame"
```

```
str(as_tibble(iris))
```

```
## tibble [150 x 5] (S3: tbl_df/tbl/data.frame)
##  $ Sepal.Length: num [1:150] 5.1 4.9 4.7 4.6 5 5.4 4.6 5 4.4 4.9 ...
##  $ Sepal.Width : num [1:150] 3.5 3 3.2 3.1 3.6 3.9 3.4 3.4 2.9 3.1 ...
##  $ Petal.Length: num [1:150] 1.4 1.4 1.3 1.5 1.4 1.7 1.4 1.5 1.4 1.5 ...
##  $ Petal.Width : num [1:150] 0.2 0.2 0.2 0.2 0.2 0.4 0.3 0.2 0.2 0.1 ...
##  $ Species     : Factor w/ 3 levels "setosa","versicolor",..: 1 1 1 1 1 1 1 1 1 1 ...
```

```
class(as_tibble(iris))
```

```
## [1] "tbl_df"     "tbl"        "data.frame"
```

- tibbles are more strict.

3. Chapter 11. Data import

How to load excel file or csv file into R?

- **readr** package in tidyverse.
  function: read_csv()
- **readxl** package for importing .xls and xlsx

```
library(readxl)

weekly_plan <-
  readxl::read_xlsx(path = "weekly time plan.xlsx") %>%
  select(-`Preparation need from instructor`)
```

How to connect to Denodo through ODBC?

After DenodoODBC is installed, you can use library **odbc** to query Denodo views.

```
library(odbc)

# setup the connection
con <- dbConnect(odbc::odbc(), "DenodoODBC", timeout = 30)

# query table using SQL
transfile <-
  dbGetQuery(con, 'SELECT * FROM bi_corporate_results."Transfile_2022_Final"') %>%
  as_tibble()
```

4. Exercise

4.1 Exercise 1.

- Read the sheet **PBI New** and **SAS New** of Excel file: **Test PBI and SAS new and endorsement 20220214.xlsx** into R.
- How many records in each tab?
- what's the min, max, and mean of column "DateDiff" in sheet **PBI New** and column "**Bus_Days_End_to_End**" in sheet **SAS New**.

4.2 Exercise 2.

- Load the Denodo view bi_crm.Transfile_Company_Table_2022 in R.
- sum of “Small Business Financial Customer Served Partnership Count”, and sum of “Medium Financial Customer Served Count”