June 19, 2023

The results below are generated from an R script.

```
# Assignment: ASSIGNMENT 2
# Name: Smith, David
# Date: 20d23-06-18

## Check your current working directory using `getwd()`
getwd()

## [1] "F:/GitLab-Projects/Bellevue/local-smith_dsc520"

## List the contents of the working directory with the `dir()` function
dir()

##  [1] "assignment_02_SmithDavid.log" "assignment_02_SmithDavid.pdf"
##  [3] "assignment_02_SmithDavid.tex" "assignments"
##  [5] "completed"                    "data"
##  [7] "LICENSE"                      "local-smith_dsc520.Rproj"
##  [9] "README.md"                    "RMarkdown.md"

## If the current directory does not contain the `data` directory, set the
## working directory to project root folder (the folder should contain the `data` directory
## Use `setwd()` if needed
setwd("F:/GitLab-Projects/Bellevue/smith-dsc520")

## Load the file `data/tidynomicon/person.csv` to `person_df1` using `read.csv`
## Examine the structure of `person_df1` using `str()`
person_df1 <- read.csv("F:/GitLab-Projects/Bellevue/smith-dsc520/data/tidynomicon/person.csv")
str(person_df1)

## 'data.frame': 5 obs. of  3 variables:
##  $ person_id    : chr  "dyer" "pb" "lake" "roe" ...
##  $ personal_name: chr  "William" "Frank" "Anderson" "Valentina" ...
##  $ family_name  : chr  "Dyer" "Pabodie" "Lake" "Roerich" ...

## R interpreted names as factors, which is not the behavior we want
## Load the same file to person_df2 using `read.csv` and setting `stringsAsFactors` to `FALSE`
## Examine the structure of `person_df2` using `str()`
person_df2 <- read.csv("F:/GitLab-Projects/Bellevue/smith-dsc520/data/tidynomicon/person.csv", stringsAs
str(person_df2)

## 'data.frame': 5 obs. of  3 variables:
##  $ person_id    : chr  "dyer" "pb" "lake" "roe" ...
##  $ personal_name: chr  "William" "Frank" "Anderson" "Valentina" ...
##  $ family_name  : chr  "Dyer" "Pabodie" "Lake" "Roerich" ...
```

```
## Read the file `data/scores.csv` to `scores_df`
## Display summary statistics using the `summary()` function
scores_df <- read.csv("F:/GitLab-Projects/Bellevue/smith-dsc520/data/scores.csv")
summary(scores_df)
```

```
##      Count           Score         Section
##  Min.   :10.00   Min.   :200.0   Length:38
##  1st Qu.:10.00   1st Qu.:300.0   Class :character
##  Median :10.00   Median :322.5   Mode  :character
##  Mean   :14.47   Mean   :317.5
##  3rd Qu.:20.00   3rd Qu.:357.5
##  Max.   :30.00   Max.   :395.0
```

```
## Load the `readxl` library
library(readxl)

## Using the excel_sheets() function from the `readxl` package,
## list the worksheets from the file `data/G04ResultsDetail2004-11-02.xls`
voter_turnout_df <- excel_sheets("F:/GitLab-Projects/Bellevue/smith-dsc520/data/G04ResultsDetail2004-11-
voter_turnout_df
```

```
##  [1] "Instructions"          "Voter Turnout"         "President"
##  [4] "House of Rep"          "Co Clerk"              "Co Reg Deeds"
##  [7] "Co Public Defender"    "Co Comm 1"             "Co Comm 3"
## [10] "Co Comm 5"             "Co Comm 7"             "St Bd of Ed 2"
## [13] "St Bd of Ed 4"         "Legislature 5"         "Legislature 7"
## [16] "Legislature 9"         "Legislature 11"        "Legislature 13"
## [19] "Legislature 23"        "Legislature 31"        "Legislature 39"
## [22] "MCC 1"                 "MCC 2"                 "MCC 3"
## [25] "MCC 4"                 "OPPD"                  "MUD"
## [28] "NRD 3"                 "NRD 5"                 "NRD 7"
## [31] "NRD 9"                 "OPS 2"                 "OPS 4"
## [34] "OPS 6"                 "OPS 8"                 "OPS 10"
## [37] "OPS 11"                "OPS 12"                "ESU 2"
## [40] "ESU 3"                 "Arlington Sch 24"      "Bennington Sch 59"
## [43] "Elkhorn Sch 10"        "Fremont Sch 1"         "Ft Calhoun Sch 3"
## [46] "Gretna Sch 37"         "Millard Sch 17"        "Ralston Sch 54"
## [49] "Valley Sch 33"         "Waterloo Sch 11"       "Bennington Mayor"
## [52] "Elkhorn Mayor"         "Valley Mayor"          "Ralston Mayor"
## [55] "Ralston Library Bd"    "Bennington City Cnc 1" "Bennington City Cnc 2"
## [58] "Elkhorn City Cnc A"    "Elkhorn City Cnc B"    "Elkhorn City Cnc C"
## [61] "Ralston City Cnc 1"    "Ralston City Cnc 2"    "Ralston City Cnc 6"
## [64] "Waterloo Bd Trustees"  "Valley City Cnc"       "Amendment 1"
## [67] "Amendment 2"           "Amendment 3"           "Amendment 4"
## [70] "Initiative 417"        "Initiative 418"        "Initiative 419"
## [73] "Initiative 420"
```

```
## Using the `read_excel` function, read the Voter Turnout sheet
## from the `data/G04ResultsDetail2004-11-02.xls`
## Assign the data to the `voter_turnout_df1`
## The header is in the second row, so make sure to skip the first row
## Examine the structure of `voter_turnout_df1` using `str()`

voter_turnout_df1 <- read_excel("F:/GitLab-Projects/Bellevue/smith-dsc520/data/G04ResultsDetail2004-11-0
str(voter_turnout_df1)
```

```
## tibble [19 x 1] (S3: tbl_df/tbl/data.frame)
##  $ Douglas County Election Commission: chr [1:19] "November 2, 2004 Presidential General Election" "(

## Using the `read_excel()` function, read the Voter Turnout sheet
## from `data/G04ResultsDetail2004-11-02.xls`
## Skip the first two rows and manually assign the columns using `col_names`
## Use the names "ward_precint", "ballots_cast", "registered_voters", "voter_turnout"
## Assign the data to the `voter_turnout_df2`
## Examine the structure of `voter_turnout_df2` using `str()`
voter_turnout_df2 <- read_excel("F:/GitLab-Projects/Bellevue/smith-dsc520/data/G04ResultsDetail2004-11-0
str(voter_turnout_df2)

## tibble [17 x 1] (S3: tbl_df/tbl/data.frame)
##  $ Official Results: chr [1:17] NA NA NA "To view the results of a particular race, click on the tab

## Load the `DBI` library
library(DBI)

## Create a database connection to `data/tidynomicon/example.db` using the dbConnect() function
## The first argument is the database driver which in this case is `RSQLite::SQLite()`
## The second argument is the path to the database file
## Assign the connection to `db` variable
db <- dbConnect(RSQLite::SQLite(),"data/tidynomicon/example.db")

## Query the Person table using the `dbGetQuery` function and the
## `SELECT * FROM PERSON;` SQL statement
## Assign the result to the `person_df` variable
## Use `head()` to look at the first few rows of the `person_df` dataframe
person_df <- dbGetQuery(db,"SELECT * FROM PERSON;")
head(person_df)

##   person_id personal_name family_name
## 1      dyer       William        Dyer
## 2        pb         Frank     Pabodie
## 3      lake      Anderson        Lake
## 4       roe     Valentina     Roerich
## 5  danforth         Frank    Danforth

## List the tables using the `dbListTables()` function
## Assign the result to the `table_names` variable
table_names <- dbListTables(db)
table_names

## [1] "Measurements" "Person"       "Site"         "Visited"

## Read all of the tables at once using the `lapply` function and assign the result to the `tables` var
## Use `table_names`, `dbReadTable`, and `conn = db` as arguments
## Print out the tables
tables <- lapply(table_names,dbReadTable,conn = db)

## Warning:  Column `reading`:  mixed type, first seen values of type real, coercing other
values of type string

tables
```

```
## [[1]]
##    visit_id person_id quantity reading
## 1       619      dyer      rad    9.82
## 2       619      dyer      sal    0.13
## 3       622      dyer      rad    7.80
## 4       622      dyer      sal    0.09
## 5       734        pb      rad    8.41
## 6       734      lake      sal    0.05
## 7       734        pb     temp  -21.50
## 8       735        pb      rad    7.22
## 9       735      <NA>      sal    0.06
## 10      735      <NA>     temp  -26.00
## 11      751        pb      rad    4.35
## 12      751        pb     temp  -18.50
## 13      751      lake      sal    0.00
## 14      752      lake      rad    2.19
## 15      752      lake      sal    0.09
## 16      752      lake     temp  -16.00
## 17      752       roe      sal   41.60
## 18      837      lake      rad    1.46
## 19      837      lake      sal    0.21
## 20      837       roe      sal   22.50
## 21      844       roe      rad   11.25
##
## [[2]]
##   person_id personal_name family_name
## 1      dyer       William        Dyer
## 2        pb         Frank     Pabodie
## 3      lake      Anderson        Lake
## 4       roe     Valentina     Roerich
## 5  danforth         Frank    Danforth
##
## [[3]]
##   site_id latitude longitude
## 1    DR-1   -49.85   -128.57
## 2    DR-3   -47.15   -126.72
## 3   MSK-4   -48.87   -123.40
##
## [[4]]
##   visit_id site_id visit_date
## 1      619    DR-1 1927-02-08
## 2      622    DR-1 1927-02-10
## 3      734    DR-3 1930-01-07
## 4      735    DR-3 1930-01-12
## 5      751    DR-3 1930-02-26
## 6      752    DR-3       <NA>
## 7      837   MSK-4 1932-01-14
## 8      844    DR-1 1932-03-22

## Use the `dbDisconnect` function to disconnect from the database
dbDisconnect(db)

## Import the `jsonlite` library
library(jsonlite)
```

```r
## Convert the scores_df dataframe to JSON using the `toJSON()` function
scores_json <- toJSON(scores_df)

## Convert the scores dataframe to JSON using the `toJSON()` function with the `pretty=TRUE` option
scores_json2 <- toJSON(scores_df,pretty=TRUE)
```

The R session information (including the OS info, R version and all packages used):

```r
sessionInfo()

## R version 4.3.1 (2023-06-16 ucrt)
## Platform: x86_64-w64-mingw32/x64 (64-bit)
## Running under: Windows 10 x64 (build 19045)
##
## Matrix products: default
##
##
## locale:
## [1] LC_COLLATE=English_United States.utf8  LC_CTYPE=English_United States.utf8
## [3] LC_MONETARY=English_United States.utf8 LC_NUMERIC=C
## [5] LC_TIME=English_United States.utf8
##
## time zone: America/New_York
## tzcode source: internal
##
## attached base packages:
## [1] stats     graphics  grDevices utils     datasets  methods   base
##
## other attached packages:
## [1] DBI_1.1.3      readxl_1.4.2   jsonlite_1.8.5
##
## loaded via a namespace (and not attached):
##  [1] utf8_1.2.3       fastmap_1.1.1    bit_4.0.5        xfun_0.39        cellranger_1.1.0
##  [6] magrittr_2.0.3   cachem_1.0.8     glue_1.6.2       blob_1.2.4       tibble_3.2.1
## [11] knitr_1.43       memoise_2.0.1    pkgconfig_2.0.3  bit64_4.0.5      lifecycle_1.0.3
## [16] tinytex_0.45     cli_3.6.1        RSQLite_2.3.1    fansi_1.0.4      vctrs_0.6.3
## [21] compiler_4.3.1   highr_0.10       rstudioapi_0.14  tools_4.3.1      evaluate_0.21
## [26] pillar_1.9.0     rlang_1.1.1

Sys.time()

## [1] "2023-06-19 09:36:26 EDT"
```