CEIE 474 Construction Computer Application and Informatics
Assignment 3

Background

Industrial construction is a construction method that involves large-scale use of offsite prefabrication and preassembly for building facilities, such as oil or gas production facilities and petroleum refineries. Pipe spool fabrication is crucial for the successful delivery of industrial projects. Pipe spool fabrication is heavily dependent on welding, which must be sampled and inspected to ensure that welding quality requirements are met. In practice, a pipe is generally specified by three parameters respectively: Nominal Pipe Size (NPS), Pipe Schedule, and Material. NPS defines a pipe's outside diameter, Pipe Schedule defines pipe wall thickness, and Materials are categorized into Material A – Plain Carbon Steel, Material B – Alloy Steel, Material C – Stainless Steel, and Material D – Others. Inspection results to welds are recorded as 0 – no inspection performed; 1– inspected and passed; and 2 – inspected and failed.

Data Explanation

In this assignment, three CSV files are provided to represent the design information, inspection information, and projects information, see Table A, B, and C below. In this assignment, you will learn how to manipulate these various data sources and how to derive the quality performance of each weld type to support prefabrication quality control processes. Tables A1, B1, and C1 are detailed descriptions of the variables in these three CSV data files.

Table A: Design information (design_info.csv)

proj_id	spool_id	weld_id	pipe_size	schedule	category
52	3494	30232	1	STD	Material A
52	3494	30233	1	STD	Material A
52	3494	30234	1	STD	Material A
52	3495	30235	6	STD	Material B
…	…	…	…	…	…

Table A1: Variables in the design information file (design_info.csv)

Variable names	Variable descriptions
proj_id	Project ID
spool_id	Spool ID
weld_id	Weld ID
pipe_size	Nominal Pipe Size (NPS)
schedule	Pipe schedule
category	Pipe material category

Table B: Inspection information (inspection_info.csv)

weld_id	welder_id	inspection_results
30232	223	1
30233	14	1
30234	14	1
30235	211	1
…	…	…

Table B1: Variables in the inspection information file (inspection_info.csv)

Variable names	Variable descriptions
weld_id	Weld ID
welder_id	Welder ID
inspection_results	Weld inspection results, 0='No inspection performed'; 1='performed inspection and pass'; 2='performed inspection and failed'

Table C: Projects information (projects.csv)

proj_id	proj_name	proj_type
52	CNRL SUG	0
53	Petro-Canada Bantrel	0
54	CNRL SUG	1
60	Petro-Canada Bantrel	0
…	…	…

Table C1: Variables in projects file (projects.csv)

Variable names	Variable descriptions
proj_id	Project ID
proj_name	Project name
proj_type	Project type, 0='indoor work'; 1='outdoor work'; 2='unknown'

Here, we will only investigate the quality performance of **indoor work** only (i.e., welds completed in pipe fabrication shops, indicated as "0" in proj_type column). A weld type is defined by the combination of 'pipe size', 'schedule', and 'category'. For example, weld type (10, STD, Material A) represents the weld type with pipe size 10, schedule STD, and material A.

You are recommended to create an ER diagram for the given tables to help you better interpret the data.

Data Manipulation Steps

** Note: For each of the following steps, suggested functions and expected table outputs are listed for reference only. Students don't need to strictly follow these steps if they can achieve the same outcomes.

Step 1:

One of the most common issues with analyzing data is to locate, extract, and combine data from various resources into the desired dataset. In this step, you are asked to combine "design information" and "inspection information" into one table. After that, remove the rows with "N. A" cell/cells from the obtained table and name the new table Table 1.

Questions:

(a) How many rows in total (i.e., total number of welds) exist in the combined dataset after deleting N.A. values?

Useful functions:

Load data: read.csv ()

Combine tables: left_join()

Remove rows with missing values: drop_na()

Step 2:

There are two main project types in the dataset: "indoor" and "outdoor". Here, we only focus on indoor work. Modify Table 1 by selecting welds that are produced under the "indoor work" project type, and name it as Table 2.

Questions:

(a) How many welds in total have been performed under the "Indoor work" project type?

Useful functions:

Combine Table 1 and table projects: left_join()

Using combined Table 1 and table projects to filter rows with "Indoor work" project type: filter()

Expected output table:

Table 2: Welds performed under "Indoor Work" project type

Project ID	Spool ID	Weld ID	Pipe size	Schedule	Category	Welder ID	Inspection results	Project name	Project type
52	3494	30232	1	STD	Material A	223	1	CNRL SUG	0
52	3494	30233	1	STD	Material A	14	1	CNRL SUG	0
…	…	…	…	…	…	…	…	…	…

Step 3:

Knowing a weld type is characterized by its NPS, pipe schedule, and material category, you are asked to count the number of weld types from the data, as well as the number of welds have been manufactured under each weld type. To do this you may start with modifying Table 2 by combing pipe attributes (NPS, pipe schedule, and material category) as one column called "weld type", name it as Table 3. Based on

Table 3, generate Table 4 with two columns named respectively, "weld type" and "number of welds". Table 3 and Table 4 must follow the same format as the given expected output tables.

Questions:

   (a) Based on the weld classification rule, how many types of pipe welds exist?

Useful functions:

   Group pipe attributes: unite ()

   Group similar weld type and calculate corresponding weld numbers: group_by() + summarize()

Expected output table:

Table 3: Welds with weld type information

Project ID	Spool ID	Weld ID	Weld type	Welder ID	Inspection results	Project name	Project type
52	3494	30232	1, STD, Material A	223	1	CNRL SUG	0
52	3494	30233	1, STD, Material A	14	1	CNRL SUG	0
…	…	…	…	…	…	…	…

Table 4: Number of welds for each weld type

Weld type	Number of welds
23, 80, Material C	13
2, XS, Material B	841
…	…

Step 4:

During the quality inspection process, not all welds are inspected and 0, 1, and 2 in the column "inspection results" represent "no inspection performed", "inspected and passed", and "inspected and failed" respectively. In this step, you are asked to extract a summary of inspection results for the weld types.

Based on Table 3 and Table 4, print a table with columns of "Weld type ID", "weld type", "number of welds", "number of inspected welds", and "number of failed welds" and name this table Table 5. This table must be ranked in a descending order based on the "number of welds". Once obtaining Table 5, print the top 20 rows into your submission.

Useful functions:

   Sort the column: order ()

   Group Table 3 by weld type and summarize the weld information: group_by() + summarize()

   Filter out non-inspected welds: filter()

Expected output table:

Table 5: Summary for the top 20 weld types

Weld type ID	weld type	number of welds	number of inspected welds	number of failed welds
1	2, XS, Material A	37042	7475	281
2	10, STD, Material A	19451	4494	196
…	…	…	…	…

Step 5:

Inspection results are useful in terms of assessing various weld types' quality performance. A high repair rate indicates bad performance whereas low repair rate indicates good performance. Using Table 5, summarize the work proportion, inspection rate, and repair rate for each of the weld types using equations provided below. Summarize these values in a table named Table 6, Table 6 must be arranged in descending order based on number of welds. Once obtaining Table 6, paste the top 20 rows into your submission.

$$\text{work proportion} = \frac{\text{number of welds}}{\text{total number of welds}}$$

$$\text{inspecction rate} = \frac{\text{number of inspected welds}}{\text{number of welds}}$$

$$\text{repair rate} = \frac{\text{number of failed welds}}{\text{number of inspected welds}}$$

Useful functions:

> *Rename columns: colnames()*
> *Extract columns from tables: table_name $ column_names*
> *Mutate columns to tables: mutate()*

Expected output table:

Table 6: Summarized work proportion, inspection rate, and repair rate for the top 20 weld types

Weld type ID	weld type	number of welds	number of inspected welds	number of failed welds	work proportion	inspection rate	repair rate
1	2, XS, Material A	37042	7475	281	0.165	0.202	0.038
2	10, STD, Material A	19451	4494	196	0.087	0.231	0.044
…	…	…	…	…	…	…	…

Marking Scheme

Question	Mark	Report	R Script
Step 1	10	Correct value of row numbers after deleting N. A (5)	Correct presentation of Table 1 using R script (5)
Step 2	10	Correct value of row numbers for indoor work (5)	Correct presentation of Table 2 using R script (5)
Step 3	20	Correct value of weld types (10)	Correct presentation of Table 3 and Table 4 using R script (10)
Step 4	30	Correct presentation of 20 weld types (15)	Correct presentation of Table 5 using R script (15)
Step 5	30	Correct values of work proportion, inspection rate, and repair rate (15)	Correct presentation of Table 6 using R script (15)
Total	100		

Notes

- This assignment is due on October 21, 2021.
- All assignments should be completed individually.
- Please upload your assignment solutions to BlackBoard by the deadline. Late submissions will be graded but with a zero mark.
- The solutions should be clearly computer-typed in a professional fashion.
- Direct any questions to the TA and instructor during the course office hours.
- Assignment files should be named as "Assignment #, Last Name, G#.pdf/r".
- You are required to submit both assignment report and R script on BlackBorad.