liFall 2019 Final Exam

QBS 181: Data Wrangling

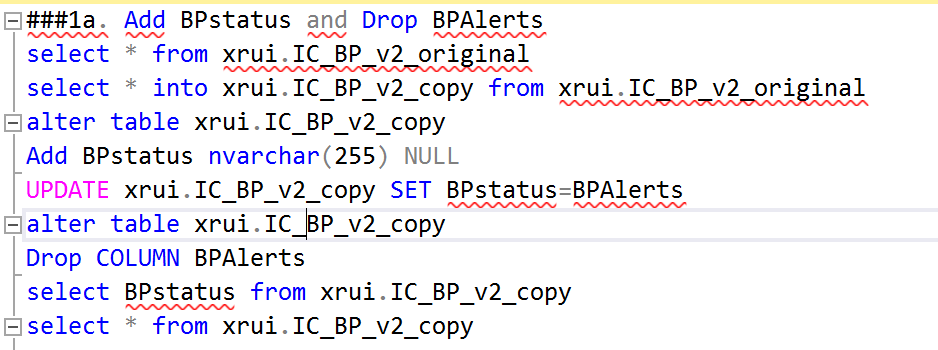
Due: 5PM, Wednesday, 11/27/2019

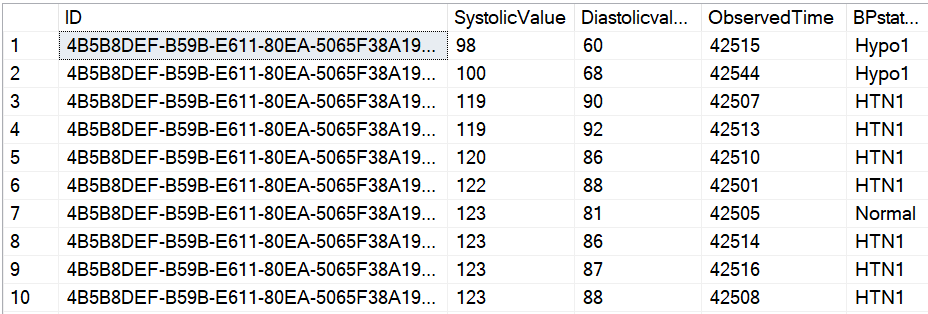
Instructions

* For this final exam, you MUST work individually, and NO collaboration is allowed. The Dartmouth Honor Principles apply ([https://www.dartmouth.edu/student-handbook/).](https://www.dartmouth.edu/student-handbook)
* For each question, print random 10 rows to show these changes. Please also include the code with comments or descriptions explaining the steps
* For Canvas submission, Please name files which makes it easily understandable. For example: *last\_first\_qbs181\_final.pdf*
* In addition to submitting the final on Canvas, you are required to deposit your code for this exam and previous assignments on GitHub (see question 4 for detail). Please also include a README.md file to help readers understand the contents of your GitHub repository
* Include the link to your GitHub repository in your report while submitting the answers.

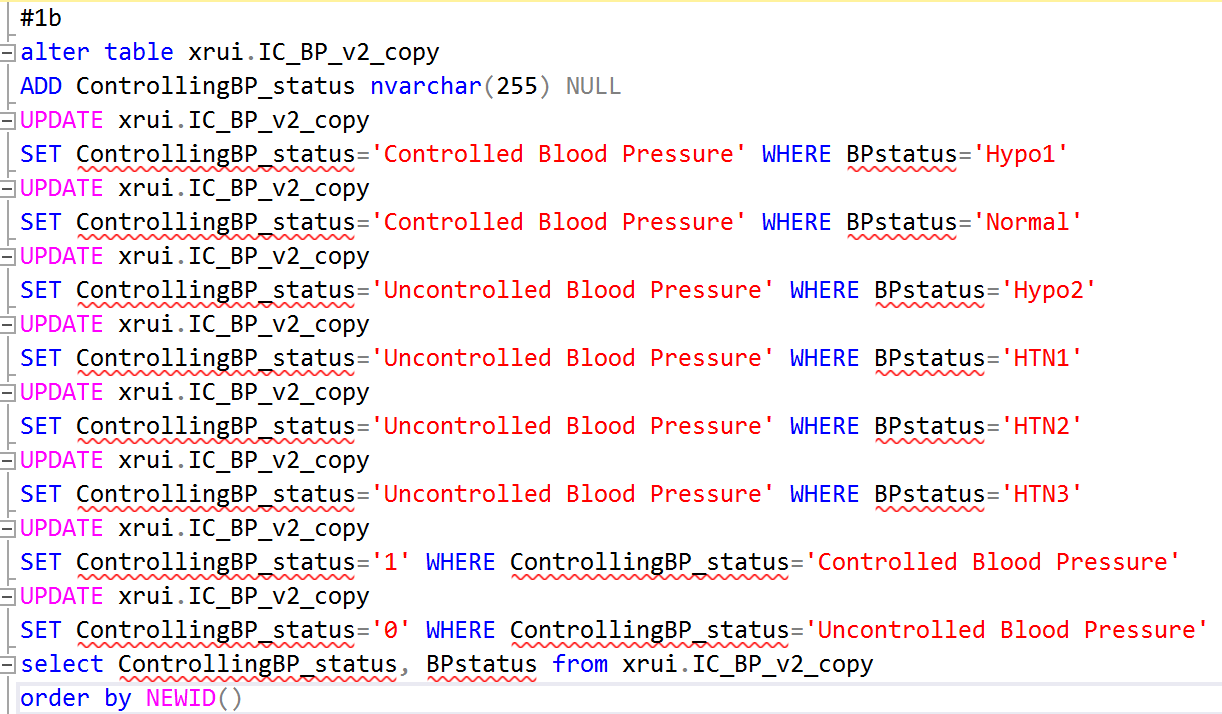
Questions

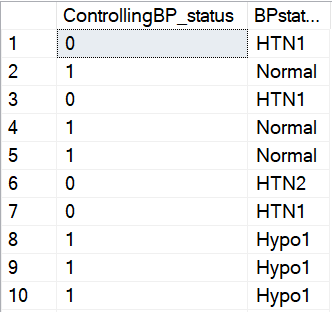
1. Consider the following blood pressure dataset (IC\_BP\_v2.csv). Perform the following operations
2. Convert BP alerts to BP statuss





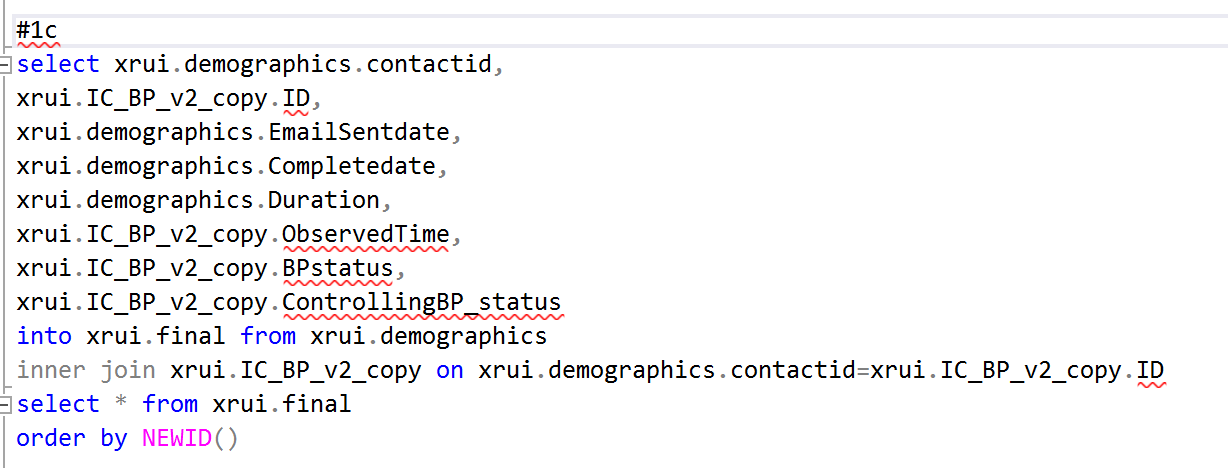
1. Define Hypotension-1 & Normal as Controlled blood pressure Hypotension-2, Hypertension-1, Hypertension-2 & Hypertension-3 as Uncontrolled blood pressure: Controlled & Uncontrolled blood pressure as 1 or 0 (Dichotomous Outcomes)

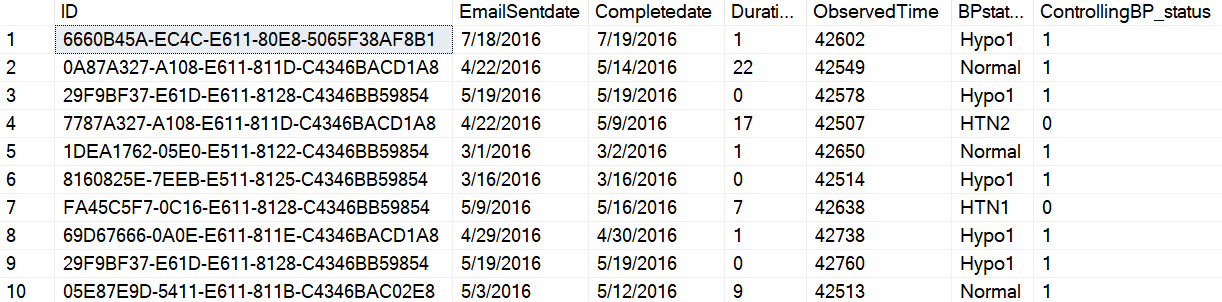




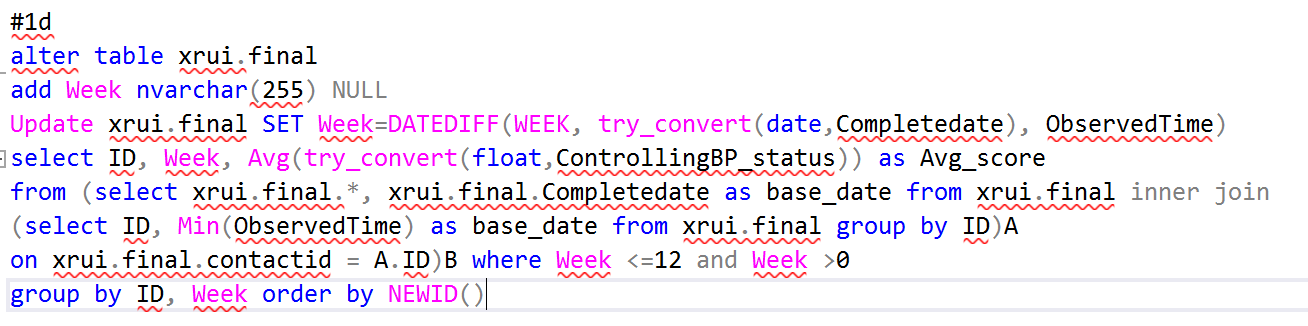
1. Merge this table with demographics (SQL table) to obtain their enrollment dates

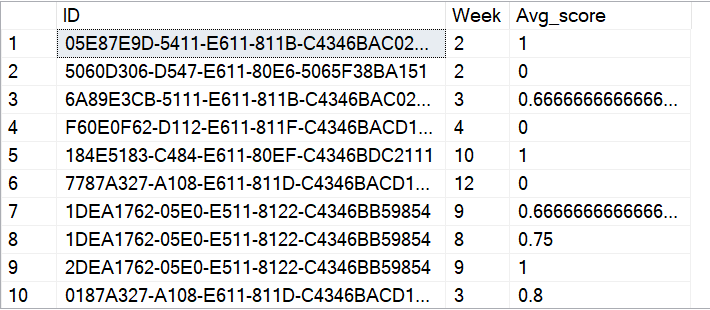
######The enrollment dates would be in column ‘Completedate’ in the table:



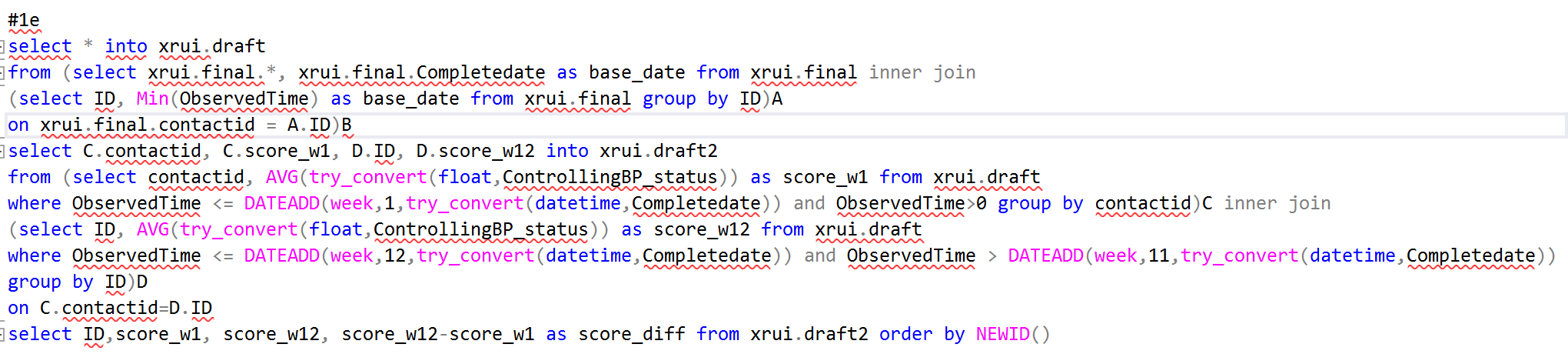


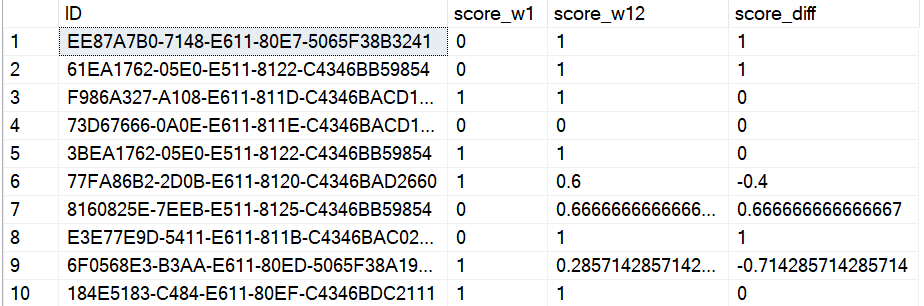
1. Create a 12-week interval of averaged scores of each customer





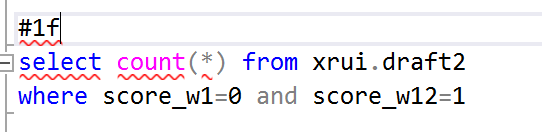
1. Compare the scores from baseline (first week) to follow-up scores (12 weeks)

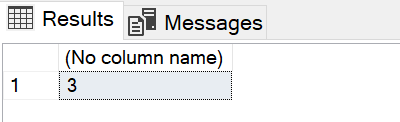




I calculated the difference using score\_w12 – score\_w1.

1. How many customers were brought from uncontrolled regime to controlled regime after 12 weeks of intervention?

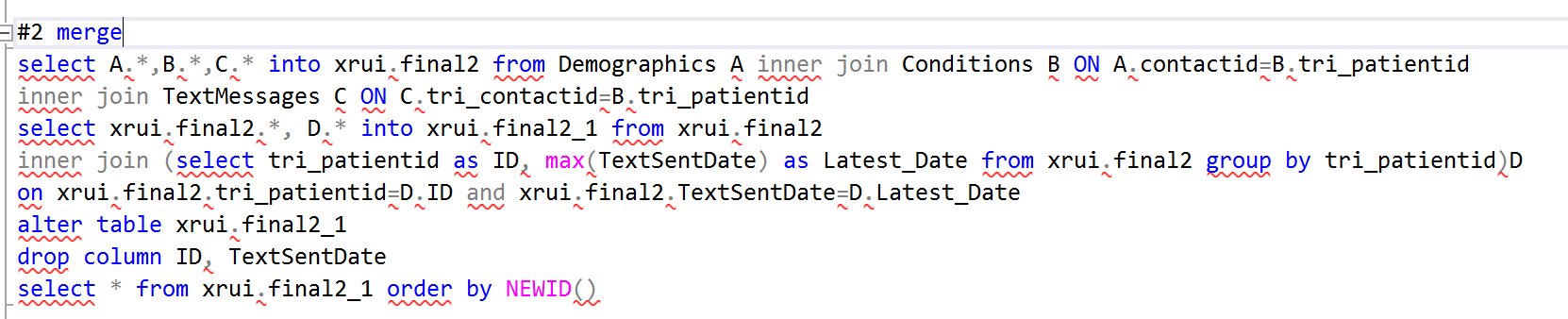


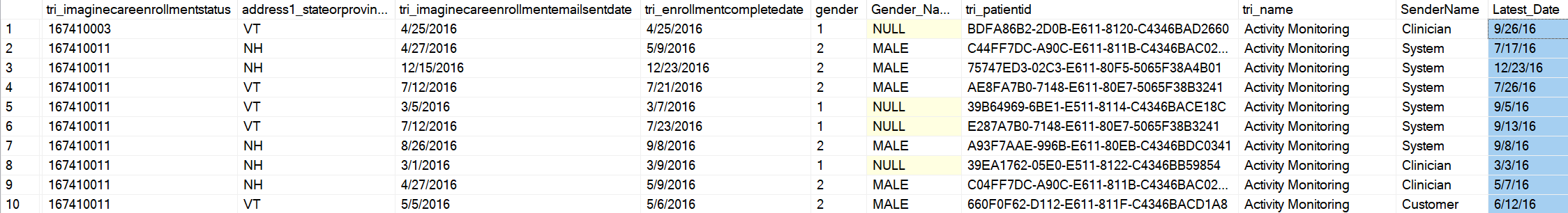


There are 3 customers.

1. Merge the tables Demographics, Conditions and TextMessages.

Obtain the final dataset such that we have 1 Row per ID by choosing on the latest date when the text was sent (if sent on multiple days)

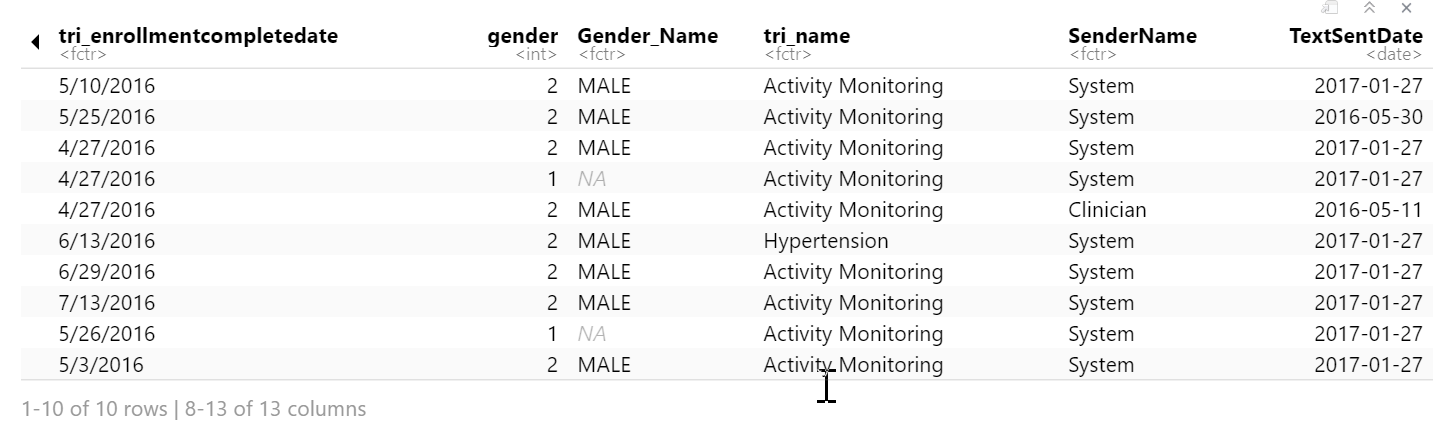




1. Repeat Question 2 in R.

Hint: You might want to use tidyr/dplyr packages





1. Set up a public GitHub repository to share your code. If you didn’t attend the Research Computing’s Git/BASH workshop, or no longer familiar with it, please use online resources (e.g. YouTube tutorials) to re-familiarize yourself with Git/GitHub.
   * Create an account on GitHub, if you haven’t done so. For this final, you are required to use the GitHub and make the repository public.
   * Create a new repository called “Data\_Wrangling\_Project\_and\_Tasks”. This repository should be a public, stand-alone repository.
   * Create sub-directories named “project1”, “project2”, “project3”, “midterm\_project”, “final\_project”, etc.
   * Copy your SQL and R code for each task into their respective repository.
     + Please include R and SQL code only.
     + Important: Do NOT include data, intermediate results, or your
   * Include a single *README.md* file for the entire repository, containing:
     + Description of the purpose of the repository.
     + A statement (1 sentence) that you have been given the permission to make your work public
     + A description of each sub-directory (you do not need to explain every file; just every task
   * Provide the link of your GitHub repository.