

Name _____ ID _____ Seat _____

King Mongkut's University of Technology Thonburi

Department of Computer Engineering

Midterm Examination 2/2017

CPE 213 Data Models

Date: March 5th, 2018

Time: 9:00 – 12:00

Instructions:

key 23/3/61

1. Carefully read the explanation in each problem and then answer each question.
 2. **Do not take the examination sheets out** of the examination room.
 3. Write your name, ID, and seat number on every page of examination sheets.
 4. Books and notes are NOT allowed to the exam room.
 5. University-certified calculator is allowed.
 5. This examination has 7 pages, 10 questions, 50 points
-

1. (6 points) Explain the definition of different types of data in a customer journey.
Give 2-3 examples of data for each type.

Name _____ ID _____ Seat _____

2. (6 points) What are the steps in data science study? Explain the required tasks in each steps in detail.

Name_____ ID_____ Seat_____

3. (3 points) Explain the difference between character and factor in R.

4. (3 points) Explain what will be the value of y for the following code. Also explain the reason why the results are like that.

```
x <- 1:6  
x + c(-1,1)  
y <- x - 1
```

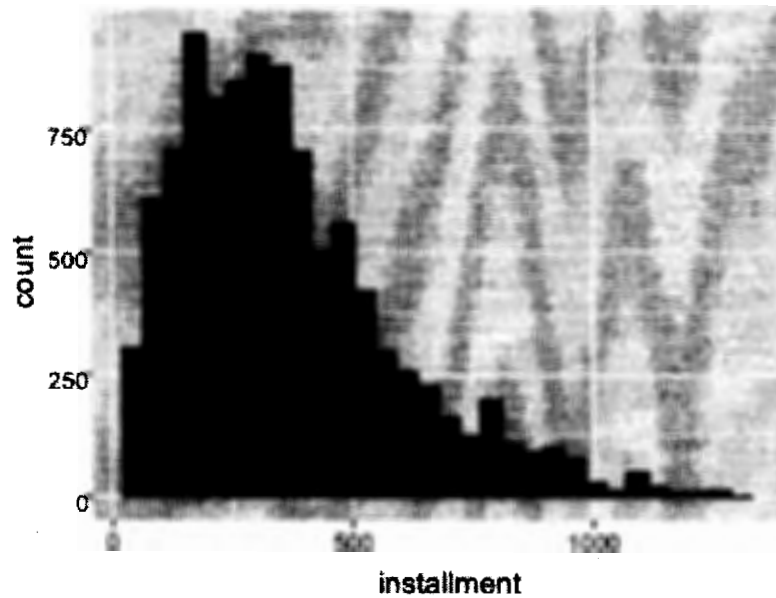
Name _____ ID _____ Seat _____

5. (4 points) Explain the functions of dplyr and their relations to the step in data preprocessing. Provides at least 4-5 explanations of functions.

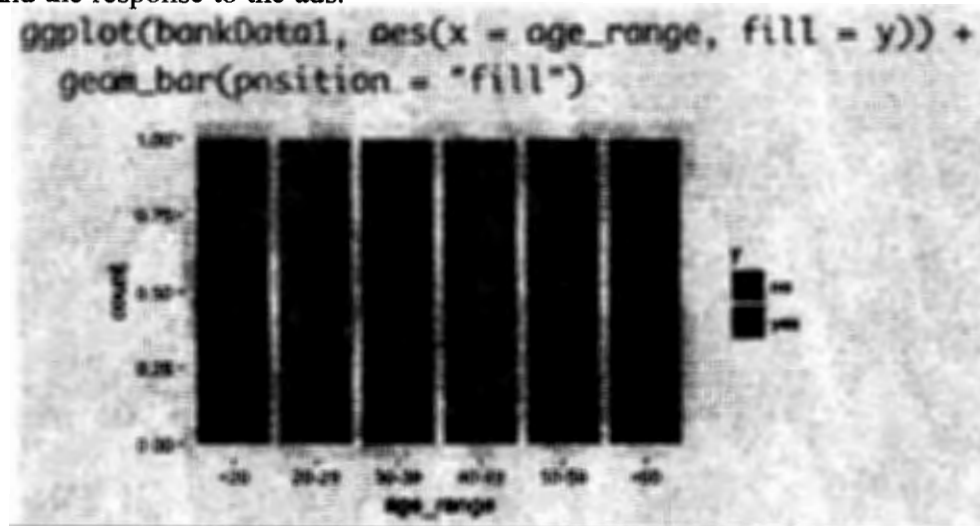
6. (5 points) From the following contingency table (row: marital, column: churn), do you think the marital status relate to the churn? Explain your reason.

	no	yes
divorced	4585	622
married	24459	2755
single	10878	1912

7. (5 points) The following graph is the histogram of “installment” (monthly payment to the debt/mortgage) of the application LendingClub. LendingClub provides peer-to-peer lending to their users which a user can submit a loan application and the application may be funded by other users. Explain, in both technical and business contexts, what you can observe in terms of distribution from this histogram?



8. (5 points) Explain what this graph tell you about the relationship of age range and the response to the ads.



9. (5 points) Explain the concepts of correlation. How does it measure the relationship between variables?

Name _____ ID _____ Seat _____

10. (8 points) Considering the social network (e.g. Facebook), if you obtain its data in forms of relationship between users (Friend), explain what you can analyze from the social network using igraph. How? Provide at least 2-3 examples.