Final Exam review part 2

Material on the Final exam:

1. Regular expressions: lectures 20, 21, 22
2. Importing CSV files into R: lecture 23
3. Unix commands (excluding R BATCH jobs): lectures 26, 27
4. sed: lecture 28
5. XML (excluding KML, HTML): lectures 29, 30, 31,
6. Xpath : lecture 32
7. Databases with dplyr: lectures 36, 9, 10 (on joins), 8 (on data verbs)

Practice problems. (Due for lab Friday May 6 as HTML file )

6. Sketch a tree for the following XML. Use circles for XML nodes and squares for content. You can ignore the attributes.­

Root Node <template>

Node

<apply>

Node

<span>

Node

<attr>

Node

<vals>

Node

<vals>

I

showBranch

Node

<attr>

Node

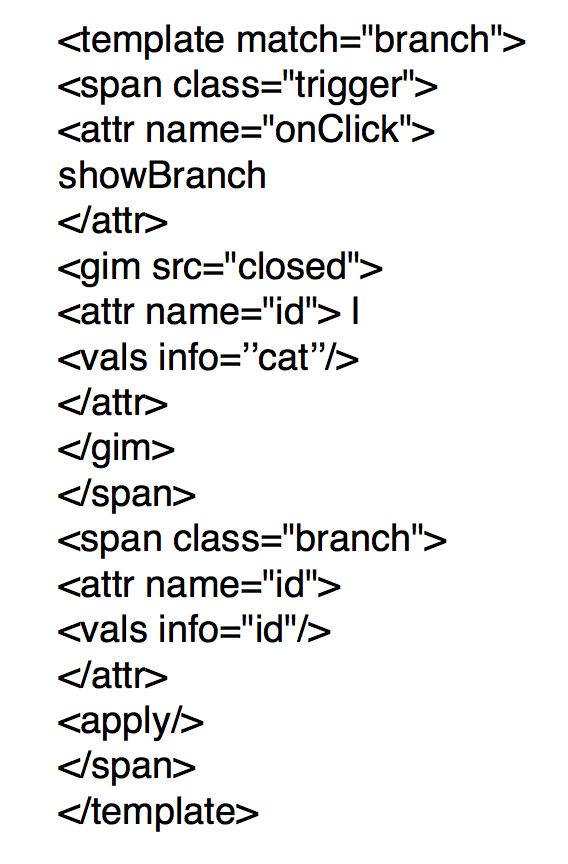
<gim>

Node

<attr>

Node

<span>



Shade the nodes in the tree that are located by the following XPath expression: //attr[@name = "id"]

Provide the return value for the following call to xpathSApply where Tree is the above XML document,

xpathSApply(Tree, ‘//attr[@name = "id"]’, xmlValue)

7. What does it mean for XML to be well-formed?  In otherwords, list the rules of an XML document (hint: see Anatomy of an XML document from lecture 29 if you get stuck)

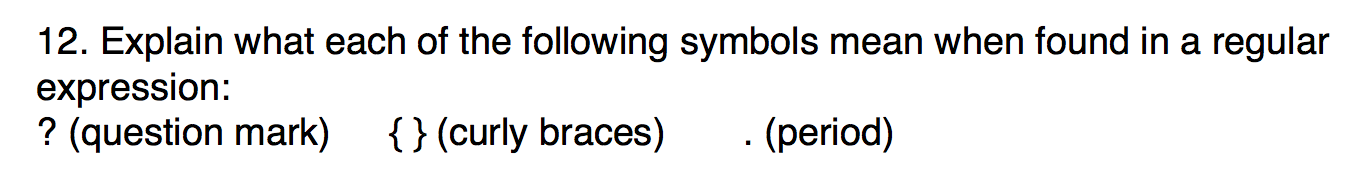
8. Which of the following represents an absolute path? (Select all that apply)

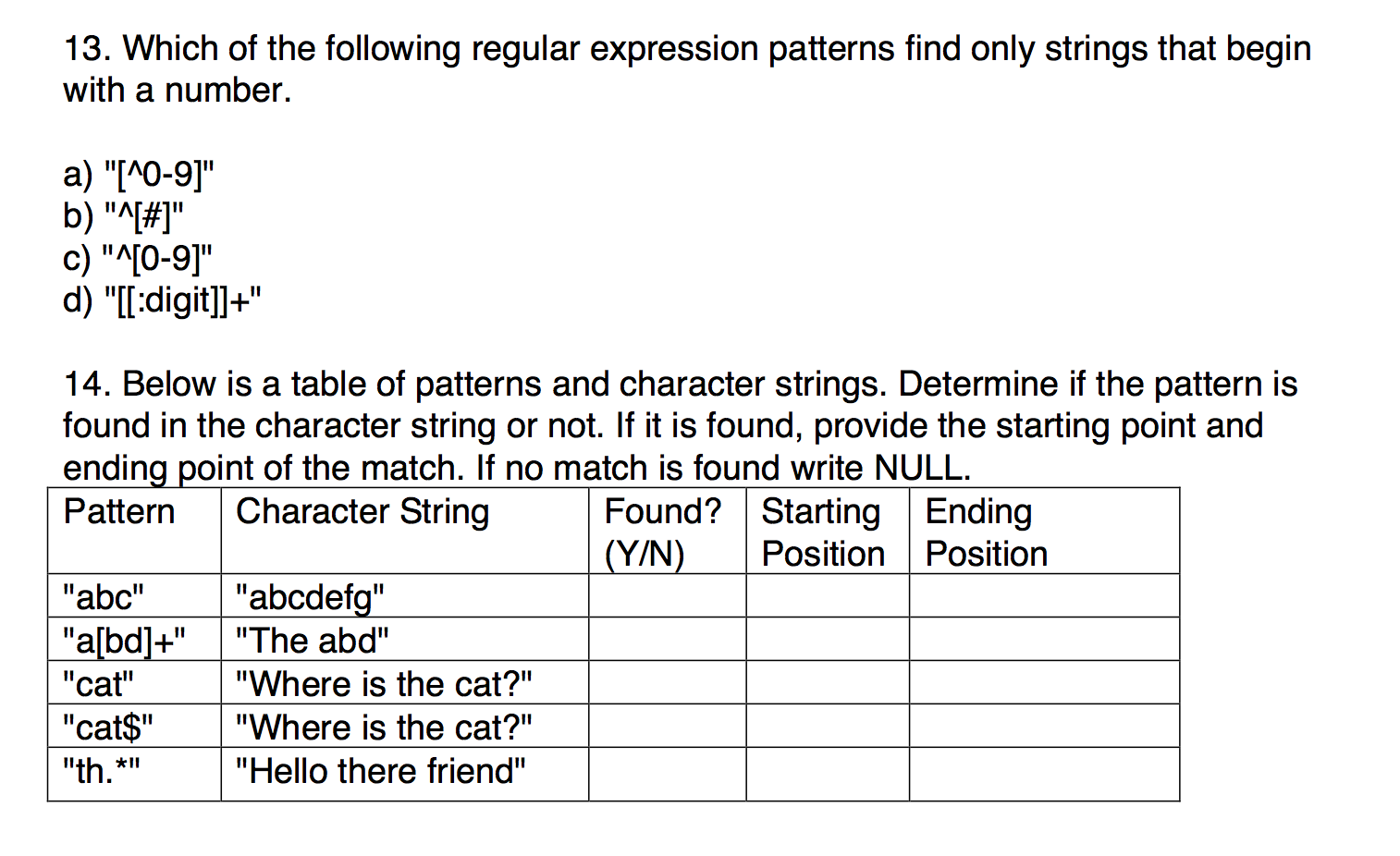
a) ../home/file.txt

b) bin/cat

c) cs2204/

d) /usr/bin/cat

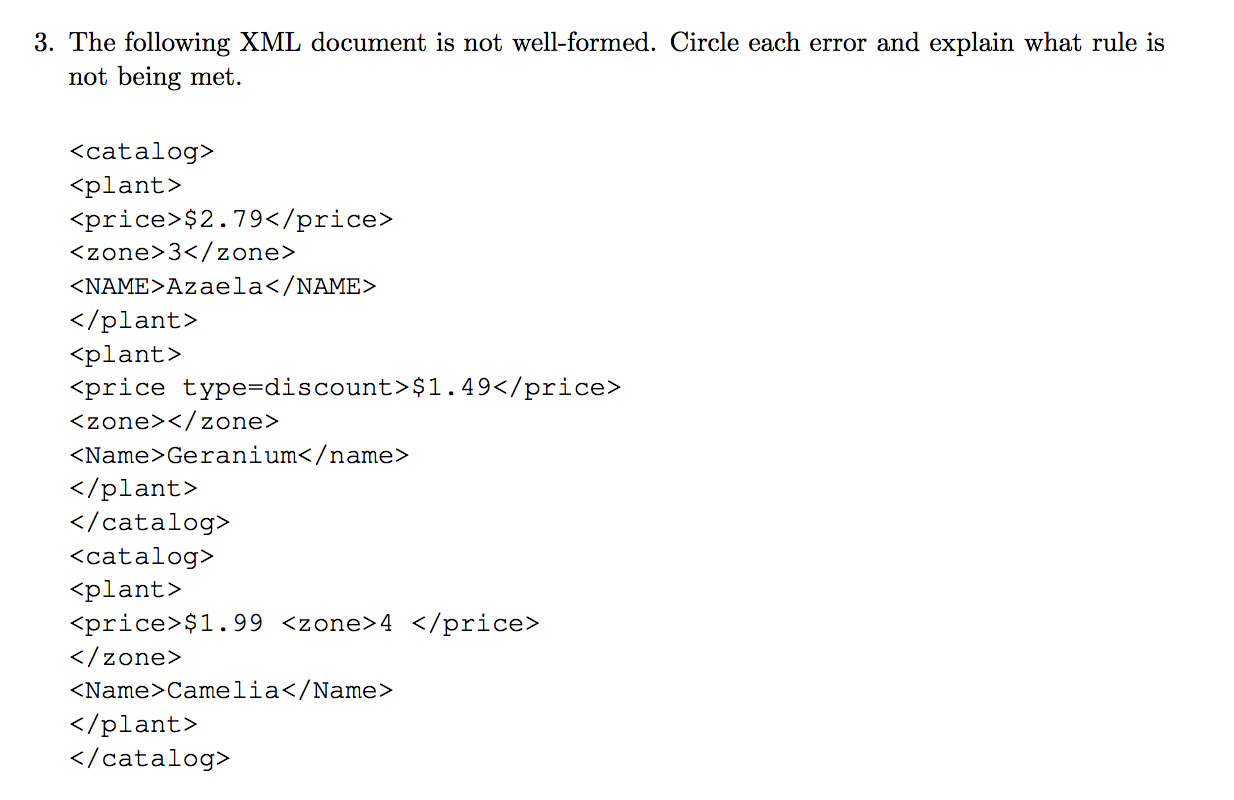


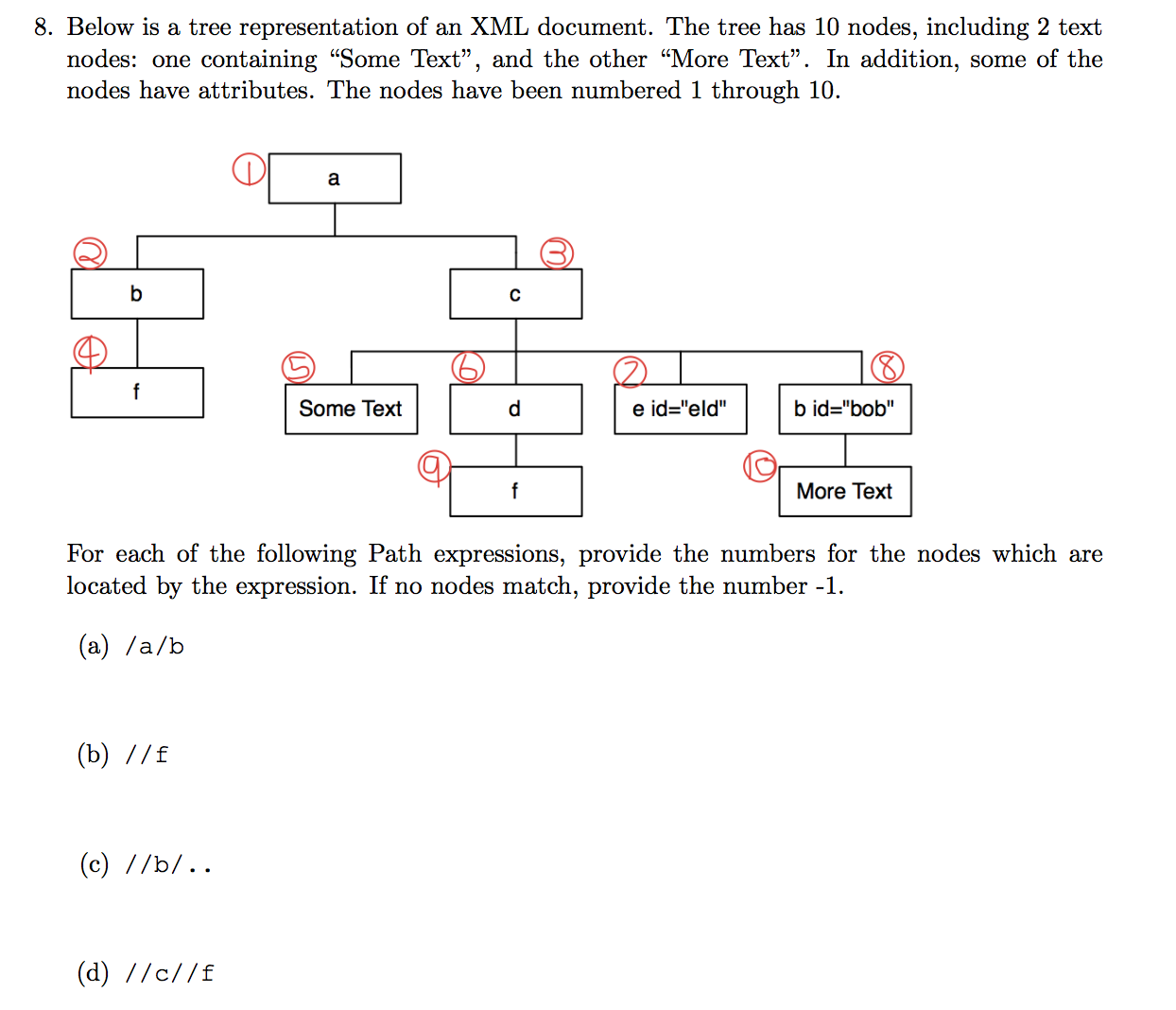


3)

a) Show all the files in the current directory that have "stat" in their name.

b) Write the results from a) to a file called “statfiles.txt”.





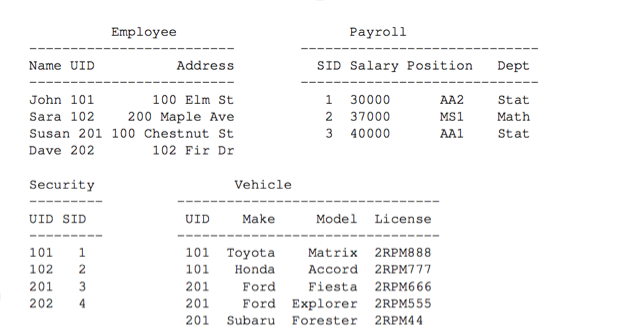
7.

Suppose that you have a relational database, called my\_db.sqlite3, on your computer. There are four tables. The data in each of the 4 tables appears below:

Employee - a record for each current and former employee with: name, address & public ID.

Security - a record containing each employee’s public ID and secure ID.

Payroll - secure ID, salary, position, and department for each current employee. Vehicle - a record for each vehicle registered with the parking office. An employee may register more than one vehicle with the office or may not have any vehicles registered.



With the dplyr package loaded you form a connection

con <- src\_sqlite(path="~/Desktop/my\_db.sqlite3")

and create tables

Employee\_sqlite <- tbl(con, "Employee")

Payroll\_sqlite <- tbl(con, "Payroll")

Security\_sqlite <- tbl(con, "Security")

Vehicle\_sqlite <- tbl(con, "Vehicle")

Using these tables and your favorite dplyr commands such as select(), mutate(), filter(), inner\_join(), group\_by, summarise(), etc provide R code to find the following.

