

homework 2

problem 1

1 . Due to recent problem of having too many files on your space in the lab share space on HPC, can you organize and concentrate your files into fewer new folders and delete all old folders and all the files that are in those old folders?

1. Here are the code snippet for one liners in terminal for question 1 answer:

creat new folder with certain names:

```
mkdir /share/jj/johndoe/newuseful1/
```

```
mkdir /share/jj/johndoe/newuseful2/
```

cd : change directory to certain folder to check if the files are important. ls -l command will display all file with the date it was created and the file's permission:

```
cd /share/jj/johndoe/olduseful1/
```

```
ls -l
```

```
cd ../olduseful2/
```

```
ls -l
```

```
cd ../olduseless1/
```

```
ls -l
```

use rm to delete any file that are not important. asterick symbol wildcard is use to replace any letters, numbers, or symbols in file name up to that point. For example, "*.json" will let you apply the command of choice to all files that ends with .json. For our example, the rm command below will remove all files that end with .json:

```
rm *.json
```

cd is to change directory and ../ indicate one level before. for example, if we are currently at /share/jj/johndoe/ directory, cd ../ will change your current directory to /share/jj/

```
cd ../
```

mv is use to move file from old directory to new directory. mv olddirectory/filename newdirectory/

```
mv /share/jj/johndoe/olduseful1/*.txt /share/jj/johndoe/newuseful1/
```

```
mv /share/jj/johndoe/olduseful2/*.txt /share/jj/johndoe/newuseful2/
```

rmdir will delete the directory and all of its subdirectories if they are all empty.

```
rmdir /share/jj/johndoe/olduseful1/ /share/jj/johndoe/olduseful2/  
/share/jj/johndoe/olduseless1/
```

problem 2

2. Can you generate a sample dataset to use for manova tutorial later (manova and its correct required dataframe type is not covered in this example)

In order to perform manova, we need dependent and independent groups. The indepdent group can be a numeric value or a text. Therefore, if you generate the sample dataset with matrix, you will not be able to create a text independent group for a numeric dependent group or vice versa. for example, if you have created dependen group using matrix and add independent group later by column:

```
ma <- matrix(1:4, nrow=2, ncol=2) ma[,3] =c("a","b")
```

or

```
ma <- matrix(c("a","b","c","d"), nrow=2, ncol=2) ma[,3] = c(1,2)
```

will give you subscript out of bounds. This is due to matrix is a homogenous type of data frame aka it can only contains either all numeric values or all text values. To trully generate the data set, we have to use dataframe in r

to generate sample data in a dataframe, you can do the following.

```
df <- data.frame(x=1:2, y=3:4)
# to check the generated data frame

df
```

```
##    x y
## 1 1 3
## 2 2 4
```

```
# assigning independent group;
df[,3] =c("a","b")
df[,4]=c(1,2)

# df now has 3 column with 1 and 2 being dependent groups for manova and 3 being indep
endent group:

df
```

```
##    x y V3 V4
## 1 1 3  a  1
## 2 2 4  b  2
```

```
# assigning name to each column :

colnames(df)=c("lenght","type","group","location")

df
```

```
##    lenght type group location
## 1      1    3    a      1
## 2      2    4    b      2
```

```
df[,1]
```

```
## [1] 1 2
```

```
df[1]
```

```
##    lenght
## 1      1
## 2      2
```

```
df[[1]]
```

```
## [1] 1 2
```

Now that you have a working data frame, you can assign a column name to each column so you can call all the values in column by \$. IF you don't assign a column name or don't want to use column name, you can simply call all value by column number aka [,1]. If you use [1,], that is to call for row 1 instead of column 1. You can also use [[1]] to call all values in column 1 (df[[1]] is similar to df[,1]). However, if you simply use df[1], it will return not only values in column but also column name, and column attribute as well. After create your own sample data, you can assign column name by

```
colnames(df)=c("lenght","type","group","location")
```

problem 3

Can you create a python script for the pipeline for the lab to use?

answer:

The student will create a pipeline.py script to do the task required. Then, to let the labmates to use, the student can do the following to set permission to use the python script. Note: dir is the directory to your pipeline.py.

```
chmod 755 dir
```

```
chmod 755 dir/pipeline.py
```