**Homework 3 Answer Sheet**

**KEY**

**Total 28 points**

## Section 1

**Question 1.1** What variables are on the x and y axis of this dataset? 1pt

X is Station, Y is Day (Note this feels weird, maybe should ask for row vs column)

**Question 1.2** How many missing values are there for each station? 1pt

Station 1, 0

Station 2, 1

Station 3, 2

Station 4, 3

**Question 1.3** If we started our analysis on day 2, only including days 2-5, how many missing values are there for each station? 1pt

Station 1, 0

Station 2, 0

Station 3, 1

Station 4, 2

**Question 1.4** Day 3? 1pt

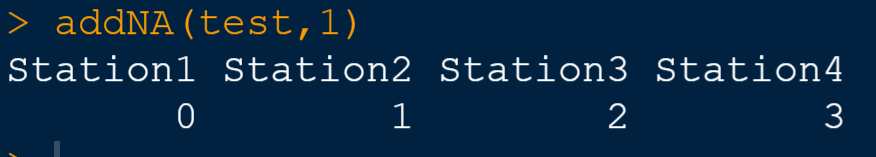
Station 1, 0

Station 2, 0

Station 3, 0

Station 4, 1

**Question 1.5** What will addNA(test,1) return? What does this mean? 2pt



Starting our analysis at day 1 will give us this many missing days.

**Question 1.6** What will addNA(test,4) return? What does this mean? 2pt



Starting our analysis at day 1 will give us this many missing days.

**Question 1.7** What are the two types of loops in R? What should we use for this purpose? 2pt

For and while loop. We should use a for loop.

**Question 1.8** What is the output of 1:dim(test)[1]? What does it mean? 2pt

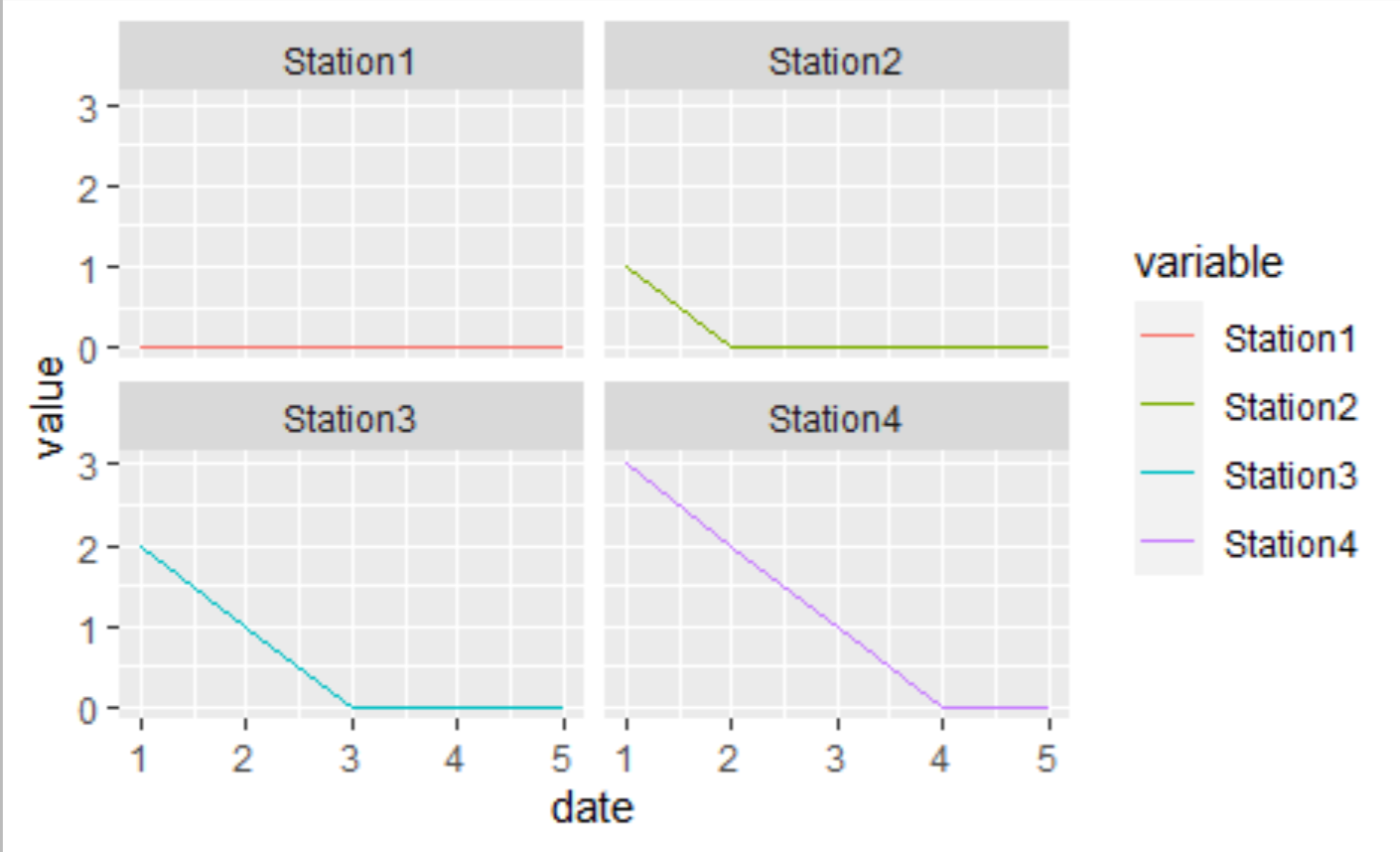
[1] 1 2 3 4 5

The range of 1 to the max row dimension of the matrix.

**Question 1.9** What will the dimensions of the output be? 1pt

The same as the input matrix.

**INCLUDE YOUR PLOT HERE!**



**Question 1.10** What do the different plots represent? 2pt

How many days are missing depending on the cutoff date.

**Question 1.11** What variables are on the x and y axis? 1pt

X axis is the cutoff date, Y is the number of missing days.

## Section 2

**Question 2.1** What are the dimensions of this data? *Hint: dim().* 1pt

[1] 18628 12

**Question 2.2** What are the column names of this data? *Hint: colnames().*1pt

[1] "date" "year" "month" "day"

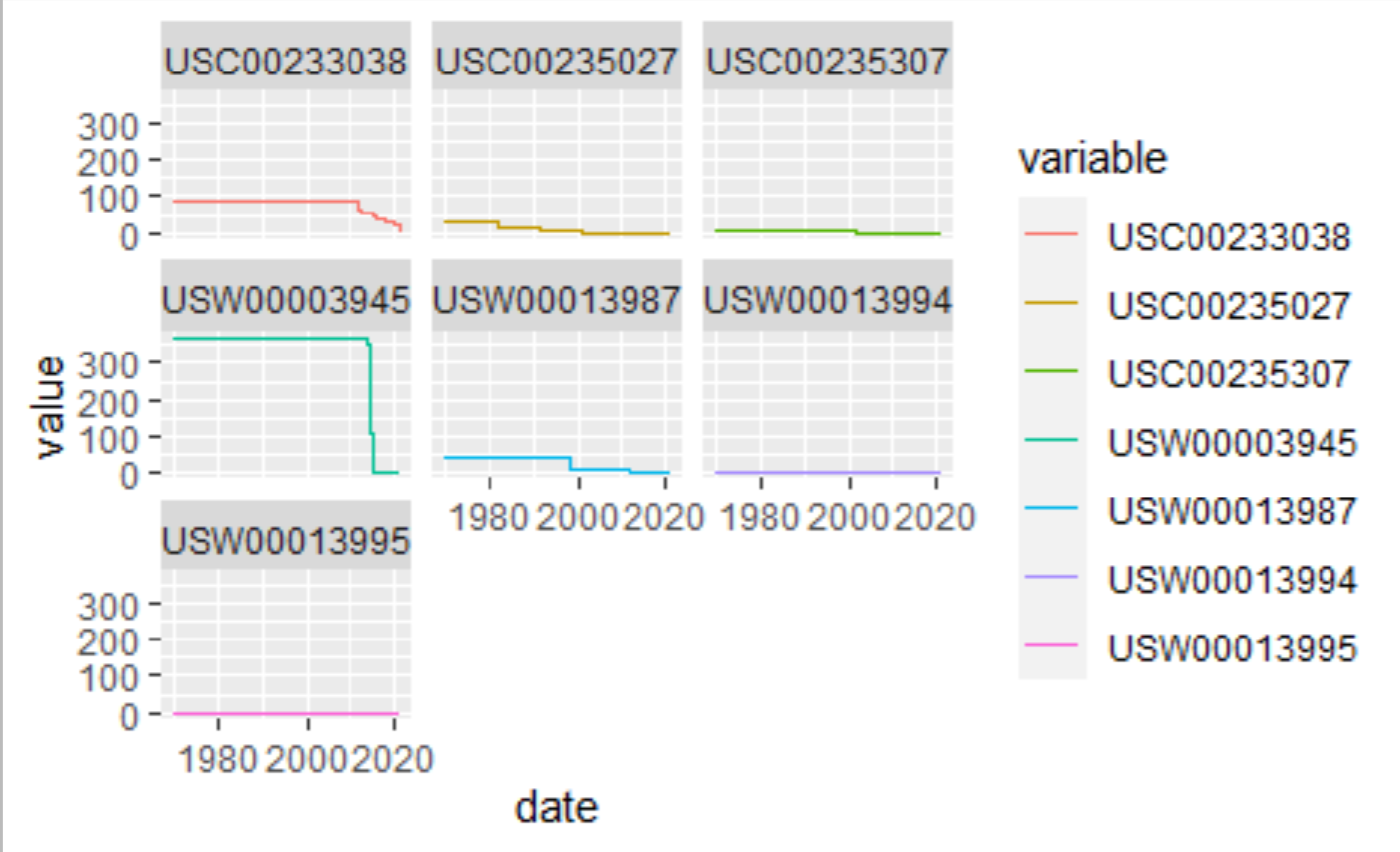
[5] "doy" "USC00233038" "USC00235027" "USC00235307"

[9] "USW00003945" "USW00013987" "USW00013994" "USW00013995"

**Question 2.3** What is the column range of the data? (Names begin with US. . . .) Use this information to make a new data frame called rain. 2pt

6:12

**INCLUDE YOUR PLOT HERE!**



**Question 2.4** Describe the missing data trends in a few of the stations. 2pt

A few have no missing dates, a few have a regular number of missing dates each year, and one has a lot of missing dates in one year.

**Question 2.5** Describe how this algorithm works. 2pt

It steps through each row and runs addNA from that row to the end of the data, counting the number of missing days.

## Section 3

**Question 3.1** What is the elapsed time of this code block? 1pt

user system elapsed

65.36 0.20 65.69

Either user or elapsed time will do!

**Question 3.2** What is the output of getDoParWorkers()? This is most likely the number of processors your computer has. 1pt

Answers vary.

**Question 3.3** What is the elapsed time of this code? Is it faster than our earlier serial code? Why is this? 2pt

user system elapsed

9.01 1.39 32.79

This is faster. The computer was able to split the work between processors.

