

Project5__2019

Elizabeth McGuckin

September 29, 2019

Worked with TA Jun and Student Natalie McGuckin

Question 1a

```
myfiles <- list.files("/class/datamine/data/airbnb/", recursive=T)
mylistings <- paste0( "/class/datamine/data/airbnb/",
                      myfiles[grepl("listings.csv",myfiles) & !grepl("listings.csv.gz",myfiles)]
)
myresults <- sapply(mylistings, read.csv)
dims <- sapply(myresults, dim)
which(dims[2,] != 16)
```

```
## /class/datamine/data/airbnb/united-kingdom/england/manchester/2016-04-10/visualisations/listings.csv
## 72
```

```
/class/datamine/data/airbnb/united-kingdom/england/manchester/2016-04-10/visualisations/listings.csv 72
```

```
myresults[[72]]$neighbourhood_group <- NA
myresults[[72]] <- myresults[[72]][,c(1:4,16,5:15)]
```

72 is missing a column

Question 1b

```
namesResult <- sapply(myresults, names)
sapply(1:16, function(x) {which(namesResult[x,] != namesResult[x])})
```

```
## [[1]]
## named integer(0)
##
## [[2]]
## named integer(0)
##
## [[3]]
## named integer(0)
##
## [[4]]
## named integer(0)
##
## [[5]]
## named integer(0)
##
## [[6]]
## named integer(0)
##
## [[7]]
```

```

## named integer(0)
##
## [[8]]
## named integer(0)
##
## [[9]]
## named integer(0)
##
## [[10]]
## named integer(0)
##
## [[11]]
## named integer(0)
##
## [[12]]
## named integer(0)
##
## [[13]]
## /class/datamine/data/airbnb/ireland/2019-06-14/visualisations/listings.csv
##                                     35
##
## [[14]]
## /class/datamine/data/airbnb/ireland/2019-06-14/visualisations/listings.csv
##                                     35
##
## [[15]]
## named integer(0)
##
## [[16]]
## named integer(0)

[[1]] named integer(0)
[[2]] named integer(0)
[[3]] named integer(0)
[[4]] named integer(0)
[[5]] named integer(0)
[[6]] named integer(0)
[[7]] named integer(0)
[[8]] named integer(0)
[[9]] named integer(0)
[[10]] named integer(0)
[[11]] named integer(0)
[[12]] named integer(0)
[[13]] /class/datamine/data/airbnb/ireland/2019-06-14/visualisations/listings.csv 35
[[14]] /class/datamine/data/airbnb/ireland/2019-06-14/visualisations/listings.csv 35
[[15]] named integer(0)
[[16]] named integer(0)

```

```
myresults[[35]] <- NULL
```

Question2

```
sapply(1:100, function(x) {is.factor(myresults[[x]]$neighbourhood)})
```

```
## [1] TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE
## [12] TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE
## [23] TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE
## [34] TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE
## [45] TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE
## [56] TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE
## [67] TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE
## [78] TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE
## [89] TRUE FALSE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE FALSE
## [100] TRUE
```

```
[1] TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE
TRUE TRUE [17] TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE
TRUE TRUE TRUE TRUE TRUE [33] TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE
TRUE TRUE TRUE TRUE TRUE TRUE TRUE [49] TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE
TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE [65] TRUE TRUE TRUE TRUE TRUE
TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE [81] TRUE TRUE TRUE TRUE TRUE
TRUE TRUE TRUE TRUE TRUE FALSE TRUE TRUE TRUE TRUE TRUE TRUE TRUE [97] TRUE TRUE
FALSE TRUE
```

```
myresults[[90]]$neighbourhood <- as.factor(myresults[[90]]$neighbourhood)
myresults[[99]]$neighbourhood <- as.factor(myresults[[99]]$neighbourhood)
```

the 6th column of these two data frames are now factors instead of integers

Question 3

```
finalData <- do.call(rbind, myresults)
```

This is now one big data frame