

exploring-redhat-data

August 14, 2016

To generate output run `rmarkdown::render('exploring_redhat_data.Rmd', output_file = 'markdown/exploring_redhat.html')`

This is an R Markdown document. Markdown is a simple formatting syntax for authoring HTML, PDF, and MS Word documents. For more details on using R Markdown see <http://rmarkdown.rstudio.com>.

read data

```
library(data.table)
activities_raw <- fread('../data/raw/act_train.csv')
```

```
##
Read 2.7% of 2197291 rows
Read 23.2% of 2197291 rows
Read 44.1% of 2197291 rows
Read 65.5% of 2197291 rows
Read 87.8% of 2197291 rows
Read 2197291 rows and 15 (of 15) columns from 0.131 GB file in 00:00:07
```

```
people_raw <- fread('../data/raw/people.csv')
```

```
summary(activities_raw)
```

```
##  people_id      activity_id      date
## Length:2197291 Length:2197291 Length:2197291
## Class :character Class :character Class :character
## Mode :character Mode :character Mode :character
##
##
## activity_category char_1      char_2
## Length:2197291 Length:2197291 Length:2197291
## Class :character Class :character Class :character
## Mode :character Mode :character Mode :character
##
##
## char_3      char_4      char_5
## Length:2197291 Length:2197291 Length:2197291
## Class :character Class :character Class :character
## Mode :character Mode :character Mode :character
##
##
## char_6      char_7      char_8
## Length:2197291 Length:2197291 Length:2197291
## Class :character Class :character Class :character
## Mode :character Mode :character Mode :character
```

```
##
##
##
##      char_9      char_10      outcome
## Length:2197291 Length:2197291 Min.   :0.000
## Class :character Class :character 1st Qu.:0.000
## Mode  :character Mode  :character Median :0.000
##                                     Mean  :0.444
##                                     3rd Qu.:1.000
##                                     Max.   :1.000
```

```
summary(people_raw)
```

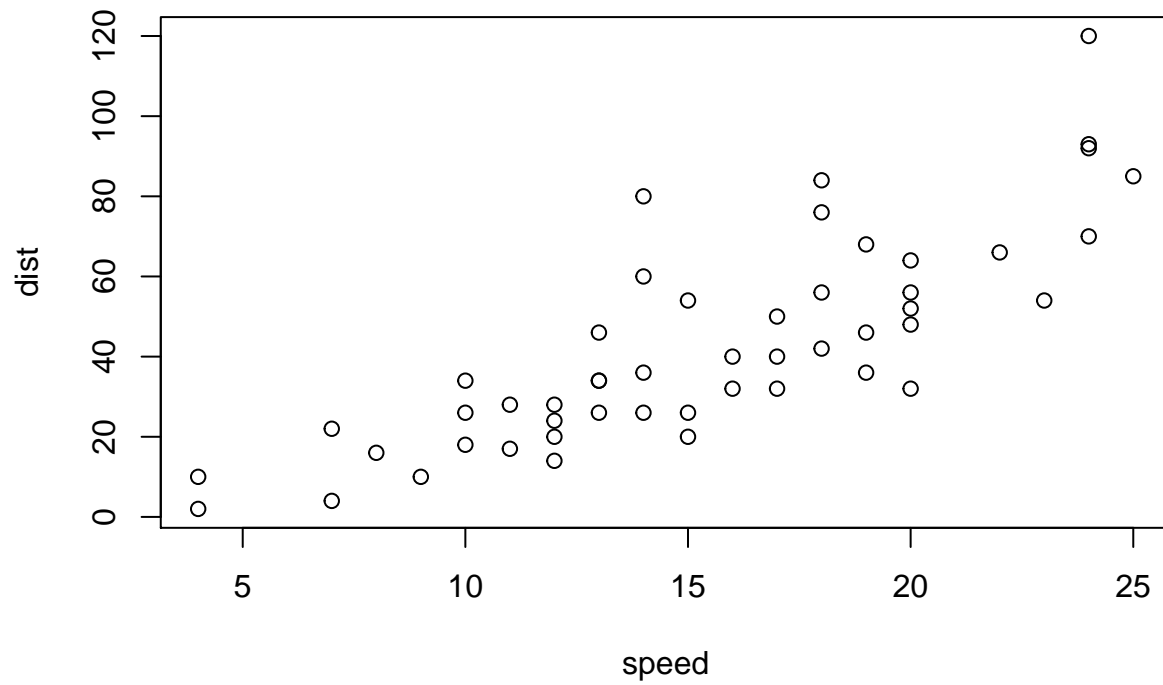
```
##  people_id      char_1      group_1
## Length:189118 Length:189118 Length:189118
## Class :character Class :character Class :character
## Mode  :character Mode  :character Mode  :character
##
##
##
##      char_2      date      char_3
## Length:189118 Length:189118 Length:189118
## Class :character Class :character Class :character
## Mode  :character Mode  :character Mode  :character
##
##
##
##      char_4      char_5      char_6
## Length:189118 Length:189118 Length:189118
## Class :character Class :character Class :character
## Mode  :character Mode  :character Mode  :character
##
##
##
##      char_7      char_8      char_9      char_10
## Length:189118 Length:189118 Length:189118 Mode :logical
## Class :character Class :character Class :character FALSE:141660
## Mode  :character Mode  :character Mode  :character TRUE :47458
##                                     NA's :0
##
##
##
##      char_11      char_12      char_13      char_14
## Mode :logical Mode :logical Mode :logical Mode :logical
## FALSE:148363 FALSE:143664 FALSE:120076 FALSE:139985
## TRUE :40755 TRUE :45454 TRUE :69042 TRUE :49133
## NA's :0 NA's :0 NA's :0 NA's :0
##
##
##
##      char_15      char_16      char_17      char_18
## Mode :logical Mode :logical Mode :logical Mode :logical
## FALSE:138148 FALSE:135772 FALSE:133903 FALSE:153635
## TRUE :50970 TRUE :53346 TRUE :55215 TRUE :35483
## NA's :0 NA's :0 NA's :0 NA's :0
##
```

```

##
##   char_19      char_20      char_21      char_22
## Mode :logical Mode :logical Mode :logical Mode :logical
## FALSE:135284  FALSE:145788  FALSE:135213  FALSE:134074
## TRUE :53834   TRUE :43330   TRUE :53905   TRUE :55044
## NA's :0       NA's :0       NA's :0       NA's :0
##
##
##   char_23      char_24      char_25      char_26
## Mode :logical Mode :logical Mode :logical Mode :logical
## FALSE:132668  FALSE:153101  FALSE:127128  FALSE:157530
## TRUE :56450   TRUE :36017   TRUE :61990   TRUE :31588
## NA's :0       NA's :0       NA's :0       NA's :0
##
##
##   char_27      char_28      char_29      char_30
## Mode :logical Mode :logical Mode :logical Mode :logical
## FALSE:144098  FALSE:134484  FALSE:157281  FALSE:149983
## TRUE :45020   TRUE :54634   TRUE :31837   TRUE :39135
## NA's :0       NA's :0       NA's :0       NA's :0
##
##
##   char_31      char_32      char_33      char_34
## Mode :logical Mode :logical Mode :logical Mode :logical
## FALSE:136433  FALSE:135237  FALSE:147920  FALSE:121701
## TRUE :52685   TRUE :53881   TRUE :41198   TRUE :67417
## NA's :0       NA's :0       NA's :0       NA's :0
##
##
##   char_35      char_36      char_37      char_38
## Mode :logical Mode :logical Mode :logical Min.   : 0.00
## FALSE:149351  FALSE:124118  FALSE:135134 1st Qu.: 10.00
## TRUE :39767   TRUE :65000   TRUE :53984  Median : 58.00
## NA's :0       NA's :0       NA's :0      Mean  : 50.33
##                                     3rd Qu.: 83.00
##                                     Max.   :100.00

```

You can also embed plots, for example:



Note that the `echo = FALSE` parameter was added to the code chunk to prevent printing of the R code that generated the plot.