importanceDistribution.R

hamze

2020-01-21

```
############################################################
############################################################
##################Importance to Prominence##################
############################################################
############################################################

#####################################################
#Installing the packages
#install.packages("classInt")
#install.packages("ggplot2")
#install.packages("RColorBrewer")
#####################################################
#set workspace to this folder
setwd("D:/Work/IJGIS/R-scripts")
#####################################################
#importing the packages
library(classInt)
```

```
## Warning: package 'classInt' was built under R version 3.5.3
```

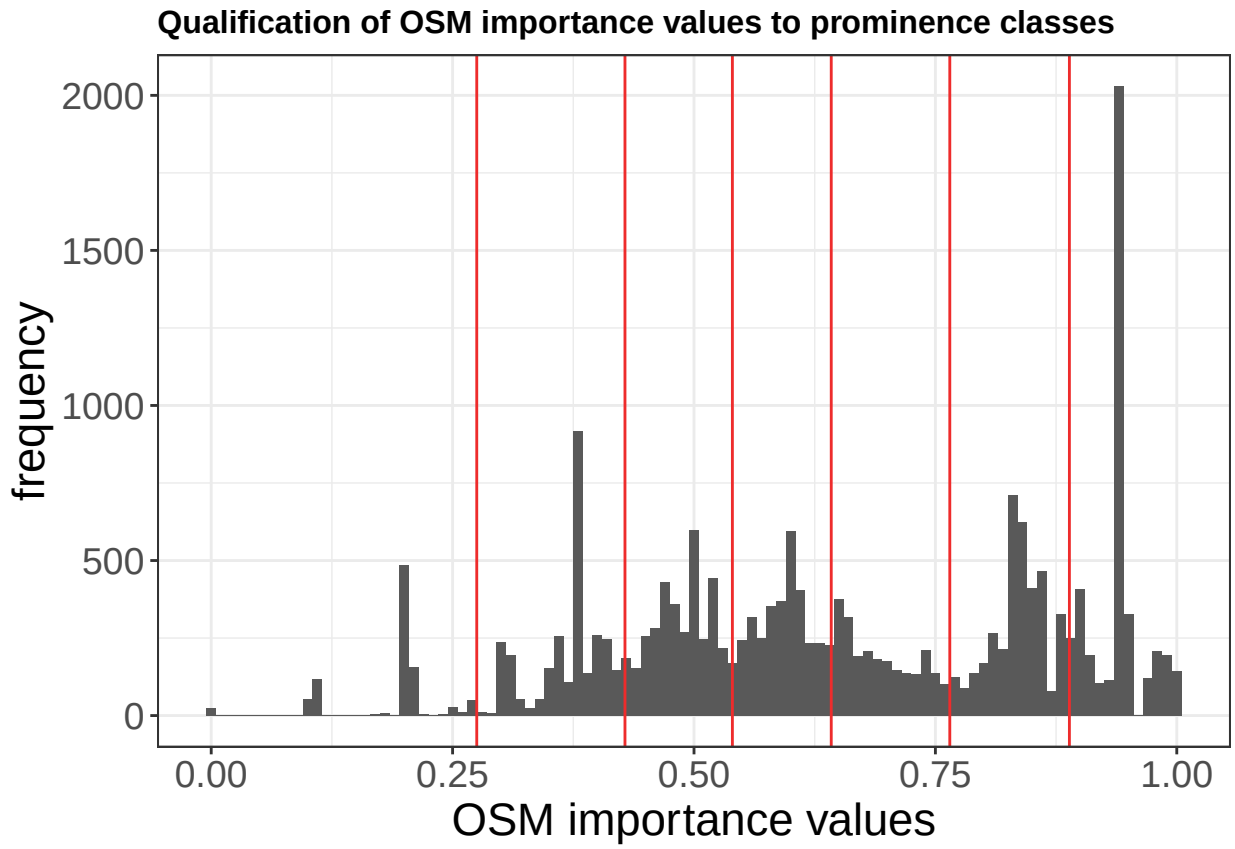
```
library(ggplot2)
```

```
## Warning: package 'ggplot2' was built under R version 3.5.3
```

```
library(RColorBrewer)


###############Reading Files###############
raw_importance <- read.table("../sequences/impotance-distr.txt", header= FALSE)
raw_importance[raw_importance > 1] <- 0.999
# set max value
raw_importance_question <- read.table("../sequences/impotance-distr-Q.txt", header= FALSE)
raw_importance_question[raw_importance_question > 1] <- 0.999
# set max value
raw_importance_answer <- read.table("../sequences/impotance-distr-A.txt", header= FALSE)
raw_importance_answer[raw_importance_answer > 1] <- 0.999
#set max value
##################NL Jenk##################
clInterval_Q <- classIntervals(raw_importance_question$V1, 7, style = "jenks")
clInterval_A <- classIntervals(raw_importance_answer$V1, 7, style = "jenks")
clInterval <- classIntervals(raw_importance$V1, 7, style = "jenks")
```

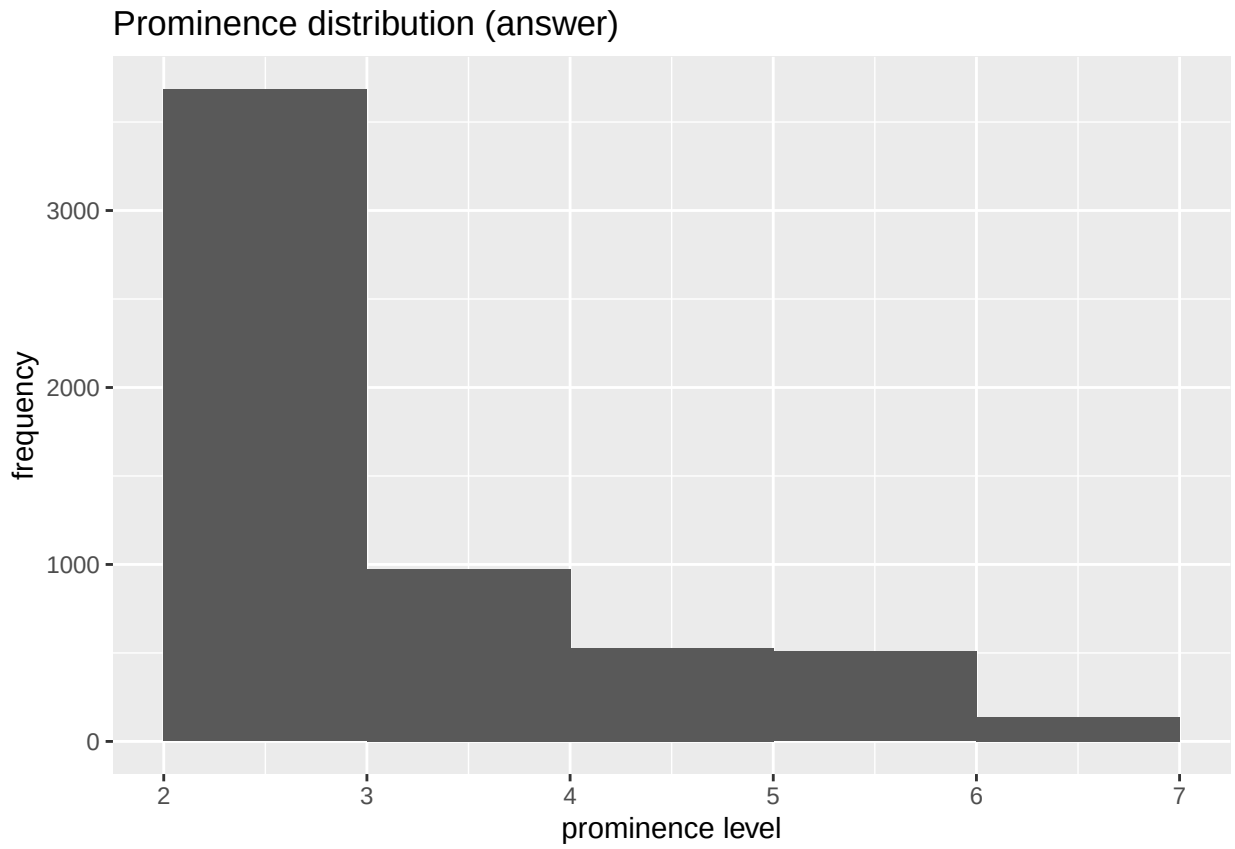
```
## Warning in classIntervals(raw_importance$V1, 7, style = "jenks"): N is large,
## and some styles will run very slowly; sampling imposed

## Use "fisher" instead of "jenks" for larger data sets
```

```
df_all <- as.data.frame(raw_importance)
df_breaks <- as.data.frame(clInterval$brks)

for (i in 1:7) {
  clInterval_Q$value[clInterval_Q$var > df_breaks[i,1]] <- i
  clInterval_A$value[clInterval_A$var > df_breaks[i,1]] <- i
}

df_answer <- as.data.frame(clInterval_A$value)
df_question <- as.data.frame(clInterval_Q$value)

#Main plot (all data!)
ggplot(df_all, aes(x=df_all$V1)) + geom_histogram(binwidth = 0.01) +
  labs(
    title="Qualification of OSM importance values to prominence classes",
    x="OSM importance values", y = "frequency") +
  geom_vline(xintercept = c(df_breaks[2:7,]), colour="firebrick2") +
  theme_bw() + theme(plot.title = element_text(color = "black", size = "12", face = "bold"),
                     text = element_text(color = "black", size=17))
```



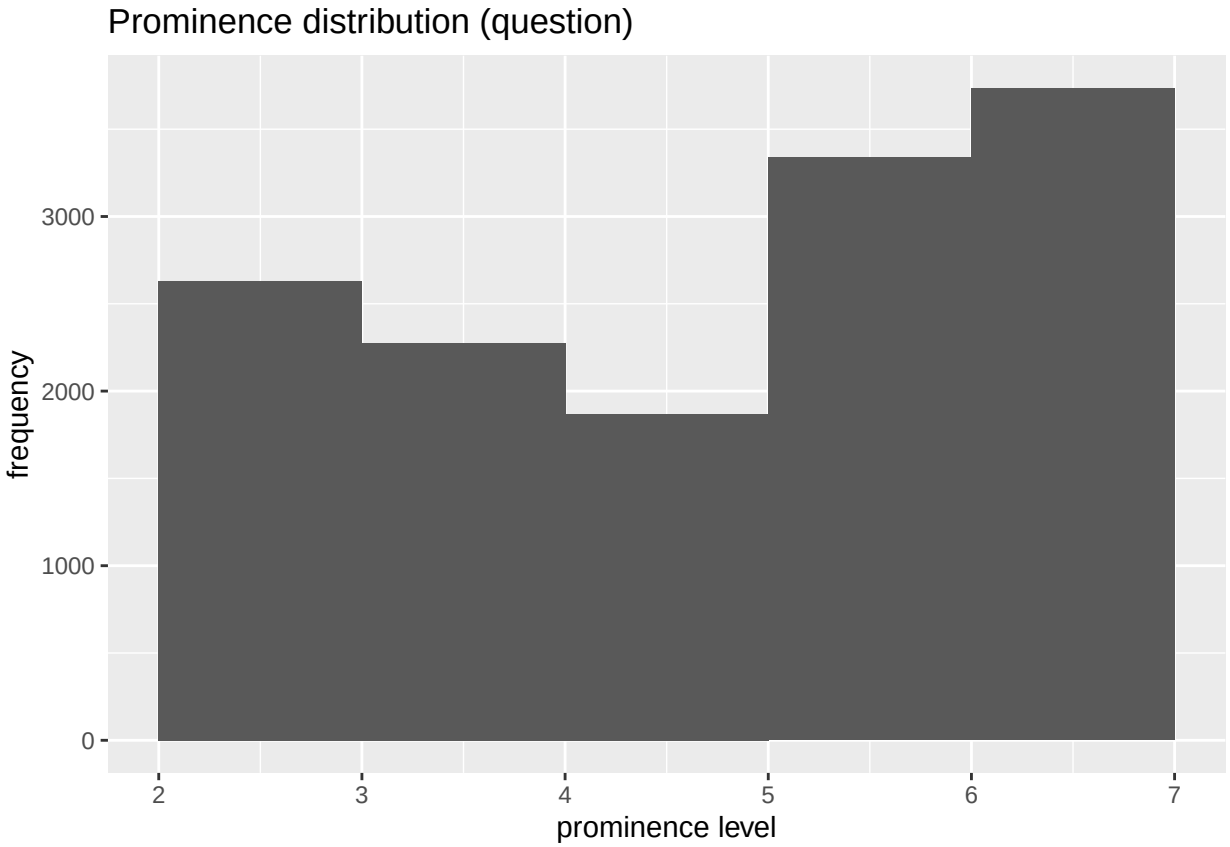
```
ggplot(df_question, aes(x=clInterval_Q$value)) +
  geom_histogram(breaks = c(2:7))+
  labs(title="Prominence distribution (answer)",x="prominence level", y = "frequency")
```

```
## Warning: Removed 9 rows containing non-finite values (stat_bin).
```




```
ggplot(df_answer, aes(x=clInterval_A$value)) +
  geom_histogram(breaks = c(2:7))+
  labs(title="Prominence distribution (question)",x="prominence level", y = "frequency")
```

```
## Warning: Removed 15 rows containing non-finite values (stat_bin).
```


Prominence distribution (question)
frequency
3000
2000
1000
0
2
3
4
5
6
7
prominence level