R Notebook

Author: Guotai Sun

criteria According to Alex, the school must be… safety – (low crime) urban – Maria wants to live the big city life, and diversity – Maria wants to study in a university with a diverse culture quality – the school should offer a decent educational program

criteria 1.Sum of all crimes from Crime_2015 data set 2.degree_urvanization = 1 (Large city) from “CollegeScorecardDataDictionary-09-12-2015” 3.’CCSIZSET’.The Carnegie Foundation classifies institutions in several ways. These data include the size and setting classification (CCSIZSET). 4.(minority_serving.historically_black, minority_serving.predominantly_black, minority_serving.annh, minority_serving.tribal, minority_serving.aanipi, minority_serving.hispanic, minority_serving.nant, =1) from “CollegeScorecardDataDictionary-09-12-2015” quality – the school should offer a decent educational program 5.degrees_awarded.highest from “CollegeScorecardDataDictionary-09-12-2015” only degrees equal and above Bachelor’s degree(>=3) 6.CURROPER (Flag for currently operating institution, 0=closed, 1=operating) from “CollegeScorecardDataDictionary-09-12-2015” 7.ADM_RATE (Admission rate) pick the top 45 or less from "CollegeScorecardDataDictionary-09-12-2015

```
#import, clean, prepare, and merge data
library(tidyverse)
```

```
## Warning: package 'tidyverse' was built under R version 4.0.5
```

```
## -- Attaching packages --------------------------------------- tidyverse 1.3.1 --
```

```
## v ggplot2 3.3.3     v purrr   0.3.4
## v tibble  3.1.4     v dplyr   1.0.7
## v tidyr   1.1.3     v stringr 1.4.0
## v readr   1.4.0     v forcats 0.5.1
```

```
## Warning: package 'tibble' was built under R version 4.0.5
```

```
## Warning: package 'tidyr' was built under R version 4.0.5
```

```
## Warning: package 'dplyr' was built under R version 4.0.5
```

```
## -- Conflicts ------------------------------------------ tidyverse_conflicts() --
## x dplyr::filter() masks stats::filter()
## x dplyr::lag()    masks stats::lag()
```

```
CollegeScorecard = read_csv("CollegeScorecard.csv")
```

```
##
## -- Column specification --------------------------------------------------------
## cols(
##   .default = col_logical(),
##   UNITID = col_double(),
##   OPEID = col_double(),
##   opeid6 = col_double(),
##   INSTNM = col_character(),
##   CITY = col_character(),
##   STABBR = col_character(),
##   ZIP = col_character(),
##   AccredAgency = col_character(),
##   INSTURL = col_character(),
##   NPCURL = col_character(),
##   HCM2 = col_double(),
##   main = col_double(),
##   NUMBRANCH = col_double(),
##   PREDDEG = col_double(),
##   HIGHDEG = col_double(),
##   CONTROL = col_double(),
##   st_fips = col_double(),
##   region = col_double(),
##   LOCALE = col_double(),
##   LATITUDE = col_double()
##   # ... with 535 more columns
## )
## i Use `spec()` for the full column specifications.
```

```
Crime = read_csv("Crime_2015.csv")
```

```
##
## -- Column specification --------------------------------------------------------
## cols(
##   MSA = col_character(),
##   ViolentCrime = col_number(),
##   Murder = col_double(),
##   Rape = col_double(),
##   Robbery = col_double(),
##   AggravatedAssault = col_double(),
##   PropertyCrime = col_number(),
##   Burglary = col_number(),
##   Theft = col_number(),
##   MotorVehicleTheft = col_double(),
##   State = col_character(),
##   City = col_character()
## )
```

```
new_CollegeScorecard <- select(CollegeScorecard, `UNITID`,`INSTNM`, `CCSIZSET`, `CURROPER`, `ADM_RATE`,`HBCU`,`PBI`, `ANNHI`,`TRIBAL`,`NANTI`,`HSI`,`AANAPII`,`SAT_AVG_ALL`,`HIGHDEG`,`LOCALE` )

new_CollegeScorecard <- mutate(new_CollegeScorecard, diversity = `HBCU`+`PBI`+ `ANNHI`+`TRIBAL`+`NANTI`+`HSI`+`AANAPII`)

Crime <- mutate(Crime, total_crime = `ViolentCrime` +`Murder`+`Rape`+`Robbery`+`AggravatedAssault`+`PropertyCrime`+`Burglary`+`Theft`+`MotorVehicleTheft`)

#safety (low crime)
new_Crime <- select(Crime, `MSA`, `total_crime` )
filter(new_Crime, `total_crime` < 3000 )
```

MSA <chr>
Bloomsburg-Berwick, PA M.S.A
Elizabethtown-Fort Knox, KY M.S.A.
Gettysburg, PA M.S.A.
Logan, UT-ID M.S.A.
Midland, MI M.S.A.
State College, PA M.S.A.
The Villages, FL M.S.A.
Wausau, WI M.S.A.
Aguadilla-Isabela, Puerto Rico M.S.A.
Arecibo, Puerto Rico M.S.A.

1-10 of 13 rows | 1-1 of 2 columns

Previous **1** 2 Next

```
#urban
filter(new_CollegeScorecard, `LOCALE` == 11 )
```

UNITID <dbl>
102553
102669
102845
103501
103644
103723
103732
103741
103787
103811

1-10 of 1,716 rows | 1-1 of 16 columns Previous **1** 2 Next

```
#diversity
filter(new_CollegeScorecard, `CCSIZSET` >= 9 )
```

UNITID <dbl>
100654
100663
100706
100724
100751
100812
100830
100858
100937
101189

1-10 of 1,433 rows | 1-1 of 16 columns Previous **1** 2 Next

```
filter(new_CollegeScorecard, `diversity` >= 1 )
```

UNITID <dbl>
100654
100724
101073
101240
101462
101569
101587
101675
101912
102030

1-10 of 333 rows | 1-1 of 16 columns Previous **1** 2 Next

```
#quality
filter(new_CollegeScorecard, `SAT_AVG_ALL` >= 900 )
```

UNITID <dbl>
100663
100706
100751
100830
100858
100937
101435
101480
101541
101587

1-10 of 1,413 rows | 1-1 of 16 columns Previous **1** 2 Next

```
filter(new_CollegeScorecard, `ADM_RATE` >= 0.5 )
```

UNITID <dbl>
100654
100663
100706
100724
100751
100830
100858
100937
101365
101435

1-10 of 1,811 rows | 1-1 of 16 columns Previous **1** 2 Next

```
filter(new_CollegeScorecard, `CURROPER` >= 1 )
```

UNITID <dbl>
100654
100663
100690
100706
100724
100751
100760
100812
100830
100858

1-10 of 7,441 rows | 1-1 of 16 columns Previous **1** 2 Next

```
filter(new_CollegeScorecard, `HIGHDEG` >= 3 )
```

UNITID <dbl>
100654
100663
100690
100706
100724
100751
100812
100830
100858
100937

1-10 of 2,952 rows | 1-1 of 16 columns Previous **1** 2 Next

```
#Add all
filter(new_CollegeScorecard, `diversity` >= 1 & `LOCALE` == 11 & `CCSIZSET` >= 9 & `HIGHDEG` >= 3
       & `CURROPER` >= 1 & `ADM_RATE` >= 0.5 & `SAT_AVG_ALL` >= 900 )
```

UNITID <dbl>
110486
110556
110617
119173
122755
125897
140553
141486
141574
160904

1-10 of 23 rows | 1-1 of 16 columns Previous **1** 2 Next

Add a new chunk by clicking the *Insert Chunk* button on the toolbar or by pressing *Ctrl+Alt+I*.

When you save the notebook, an HTML file containing the code and output will be saved alongside it (click the *Preview* button or press *Ctrl+Shift+K* to preview the HTML file).

The preview shows you a rendered HTML copy of the contents of the editor. Consequently, unlike *Knit*, *Preview* does not run any R code chunks. Instead, the output of the chunk when it was last run in the editor is displayed.