

```

1  #This script is for working with the full Basic Skills revenue and expenditure data
2  #this encompasses both compensatory revenue and English Learner revenue
3  #We received data on the revenue side (the "basicskills" file) on April 18, 2019 from
   MDE
4  #The UFARS file we had received earlier and that shows all the spending (program code
   219 is spending for English Learners)
5
6
7
8
9  #install.packages("pastecs")
10 library(pastecs)
11
12 library(readr) #importing csv files
13 library(dplyr) #general analysis
14 library(ggplot2) #making charts
15 library(lubridate) #date functions
16 library(reshape2) #use this for melt function to create one record for each team
17 library(tidyr)
18 library(janitor) #use this for doing crosstabs
19 library(scales) #needed for stacked bar chart axis labels
20 library(knitr) #needed for making tables in markdown page
21 library(htmltools) #this is needed for Rstudio to display kable and other html code
22 library(rmarkdown)
23 library(kableExtra)
24 library(ggthemes)
25 library(stringr)
26 library(RMySQL)
27 library(readxl) #for importing Excel files
28 library(DT) #needed for making searchable sortable data table
29 library(waffle)
30 library(foreign) #for importing SPSS files
31 library(jsonlite) #for exporting JSON
32 library(car)
33 library(aws.s3) #for loading to AWS server
34 options(scipen=999)
35 library(scales)
36
37
38 # traditional districts -----
39 #this is the primary file MDE gave us showing total Basic Skills amounts for each
   district
40 #But it only includes the traditional districts (types 1 and 3)
41
42 #it's set up really wide with all the years going across, with multiple values for each
   year
43 #so this next series of code pulls it in and then rearranges it
44 basicskills <- read_csv('./data/basicskills_revenue_import.csv',
   col_types=cols(`District Number`=col_character(), `District Type`=col_character()))%>%
45   clean_names() %>% mutate(districtid=paste(district_number, district_type, '000',
   sep='-'))
46
47 #Leave out a couple columns we don't need
48 basicskills2 <- basicskills %>% select(-district_number, -district_type)
49
50 #normalize the data using melt() function

```

```

51 basicskills3 <- melt(basicskills2, id=c("districtid", "district"))
52
53 #add some new columns
54 basicskills3 <- basicskills3 %>% mutate(datayr=substr(variable, 2, 6),
55                                     yr=
56                                     as.numeric(paste('20',substr(variable,5,6), sep='')),
57                                     type=substr(variable, 8,100))
58
59
60 #this turns it back into wide table
61 #also eliminates districts that had $0 basic skills revenue in a given year
62 #these appear to all be districts that consolidated or were somehow closed
63 #but had money in earlier years
64 basicskills_final <- dcast(basicskills3 %>%
65                             select(districtid, district, yr, type, value), districtid
66                             + yr + district ~ type) %>%
67                             filter(total_basic_skills_revenue>0)
68
69 #MDE failed to include in that file some of the "extra" compensatory
70 #money that goes to some districts
71 #pilot money is for a handful of suburban districts (plus Rochester)
72 #The one-time is for only one year in the past
73 #early learning is kind of sporadic
74
75 #additional compensatory -- pilot, one-time and early learning
76 additional_comp <- read_xlsx('./data/Additional district level compensatory
77 data.xlsx', sheet='Compensatory Revenue Values', range='B4:K6611') %>%
78     clean_names() %>% mutate(yr=as.integer(str_sub(year,4,6))+2000,
79                             districtid=paste(district_number_type, '000', sep='-'))
80
81 #merge the additional comp fields into basicskills_final
82 basicskills_final <- left_join(basicskills_final, additional_comp %>%
83                                 select(districtid, yr, compensatory_one_time,
84                                     x1st_year_vpk_srp_compensatory, compensatory_pilot, total_compensatory),
85                                 by=c("districtid"="districtid", "yr"="yr"))
86
87 #get rid of null values
88 basicskills_final$el_revenue[is.na(basicskills_final$el_revenue)] <- 0
89 basicskills_final$el_concentration_revenue[is.na(basicskills_final$el_concentration_rev
90 enue)] <- 0
91 basicskills_final$total_basic_skills_revenue[is.na(basicskills_final$total_basic_skills
92 _revenue)] <- 0
93 basicskills_final$total_compensatory_revenue[is.na(basicskills_final$total_compensatory
94 _revenue)] <- 0
95 basicskills_final$compensatory_one_time[is.na(basicskills_final$compensatory_one_time)]
96 <- 0
97 basicskills_final$x1st_year_vpk_srp_compensatory[is.na(basicskills_final$x1st_year_vpk
98 srp_compensatory)] <- 0
99 basicskills_final$compensatory_pilot[is.na(basicskills_final$compensatory_pilot)] <- 0
100 basicskills_final$total_compensatory[is.na(basicskills_final$total_compensatory)] <- 0
101
102 #the column called total_compensatory_revenue is missing the pilot money

```

```

98 #and some others
99 #so use the one called total_compensatory
100
101
102 # charter schools -----
103
104
105
106 #Now we need to pull compensatory revenue for charter schools
107
108 #this is revenue by building for all district types (the first file that MDE sent us)
109 #this may not include the pilot money (not sure)
110 #need to use this to pull charter school compensatory revenue for the online chart
111
112 revenue <- read_csv('./data/compensatory_revenue_bysite_06_18.csv') %>%
113   clean_names() %>%
114   mutate(schoolid=paste(district_number, district_type, site_number, sep="-"),
115          yr=as.integer(str_sub(year,4,6))+2000,
116          districtid=paste(str_sub(schoolid,1,7), '000', sep="-"))
117
118
119
120 #grab only the district type 2 and 7 compensatory revenue - summarized to district
  level
121 comp_rev_charters <- revenue %>%
122   filter(district_type=='02' | district_type=='07') %>%
123   group_by(districtid, district_name, yr) %>%
124   summarize(comp_rev_total = sum(revenue))
125
126
127 #this is EL revenue for charter schools - district types 2 and 7
128 #sent by MDE as a separate file
129 el_rev_charters <- read_csv('./data/LEPTypes2and7.csv',
130   col_types=cols(dst_num=col_character(), dst_tye=col_character(),
131
132     lep_rev=col_double(), lep_cnc_rev=col_double())) %>%
133   clean_names() %>%
134   mutate(districtid = paste(dst_num, dst_tye, '000', sep="-"),
135          total_el = lep_rev+lep_cnc_rev,
136          yr=as.integer(str_sub(dat_yer,4,6))+2000)
137
138 #create file that has basic skills revenue totals for charter schools
139 charters_rev <- left_join(comp_rev_charters, el_rev_charters, by=c("yr"="yr",
140   "districtid"="districtid")) %>%
141   mutate(basicskills_total = total_el + comp_rev_total)
142
143
144 # for online -----
145
146
147
148 #This next set of code creates a file for a table to go with the story online
149 #it has all districts for the 17-18 school year
150 #showing compensatory revenue, EL revenue and a total basic skills amount

```

```

151
152 #first need to pull the right fields for the traditional districts from
basicskills_final
153
154 foronline <- basicskills_final %>% filter(yr==2018) %>%
155   mutate(district=toupper(district), el_total = el_revenue+el_concentration_revenue,
156     pilot = case_when(compensatory_pilot>0~'y', TRUE~'n'),
157     basicskills_total = total_compensatory+el_total) %>%
158   select(districtid, district, pilot, total_compensatory, el_total, basicskills_total)
159
160
161 #next pull out the fields we need for charter schools
162 charters_foronline <- charters_rev %>%
163   filter(yr==2018) %>%
164   mutate(district=toupper(district_name), pilot='n') %>%
165   rename(total_compensatory=comp_rev_total,
166     el_total=total_el)
167 select(districtid, district,comp_rev_total, total_el, basicskills_total)
168
169
170
171 #append the traditional schools file and charter schools file together using
bind_rows()
172 foronline <- bind_rows(foronline, charters_foronline)
173
174 #spit out a csv file to use in DataWrapper
175 write.csv(foronline, './output/district_totals_2018_foronline.csv', row.names=FALSE)
176
177
178
179
180 # ANALYSIS -----
181
182
183 #this shows the total basic skills money going to charters each year
184 charters_rev %>%
185   filter(dst_tye=='07') %>%
186   group_by(yr) %>% summarise(count=n(), tot= sum(basicskills_total), el =
187     sum(total_el), comp=sum(comp_rev_total))
188
189
190
191
192
193 # import UFARS spending data -----
194 --
195 #this shows how Basic Skills money was spent for 2005-06 through 2017-18
196 ufars06_18 <- read_csv('./data/ufars06_18.csv',
197   col_types=cols(.default=col_character(),
198   tot_amt=col_double()))%>%
199   rename(datayear=dat_yer,districtnum=dst_num, disttype=dst_tye,fund=fun_num,
200   organization=ogz_num,
201   program=prg_num, finance=fna_num,object=obj_num,course=crs_num,
202   schoolclass=unt_cls)
203

```

```

201 #this is the list of codes (this gets used in the richfield.rmd)
202 codes <- read_excel("../data/UFARS/09-ListofCodes 2019.1.xlsx", sheet="CODES",
203 range="A1:D730")
204
205 # import from mysql -----
206 #this imports names and other info on the districts
207
208
209 con <- dbConnect(RMySQL::MySQL(), host = Sys.getenv("host"), dbname="Schools",user=
210 Sys.getenv("userid"), password=Sys.getenv("pwd"))
211
212 #List the tables in the database we've connected to
213 #dbListTables(con)
214
215 #List the fields in the table; change "mytablename" to the name of the table you're
216 trying to connect to
217 #dbListFields(con, 'mytablename')
218
219 #Pull DistrictList table
220 data1 <- dbSendQuery(con, "select * from DistrictList")
221
222 #assign it to a new data frame
223 district_list <- fetch(data1, n=-1)
224
225 dbClearResult(data1)
226
227 #disconnect connection
228 dbDisconnect(con)
229 rm(data1)
230
231
232 #clean up district_list data frame
233 district_list <- district_list %>% clean_names() %>% rename(district_name=organization)
234
235
236 #add some fields to the ufars data
237 #need to limit to finance code 317 (basic skills) because they accidentally gave us
238 some bad records
239 ufars06_18 <- ufars06_18 %>%
240   filter(finance=='317') %>%
241   mutate(schoolid=paste(districtnum, disttype, organization, sep="-"),
242     yr=as.integer(str_sub(datayear,4,6))+2000,
243     districtid=paste(str_sub(schoolid,1,7),'000',sep="-"))
244
245 #split UFARS into two files
246 #note we're excluding the 2005-06 data from UFARS cause we don't have matching revenue
247 data
248 #first one is for English Learner spending under Basic Skills
249 el_spent <- ufars06_18 %>% filter(disttype=='01' | disttype=='03', program=='219',
250 yr>2006) %>%
251   group_by(yr, districtid) %>% summarise(el_spent= sum(tot_amt))
252
253 #second one is for all non-EL spending under Basic Skills

```

```

252 comp_spent <- ufars06_18 %>% filter(disttype=='01' | disttype=='03', program!='219',
yr>2006) %>%
253   group_by(yr, districtid) %>% summarise(comp_spent= sum(tot_amt))
254
255
256
257
258
259 #Start putting everything together
260
261
262 #Join what we have so far with compensatory spending
263 match <- full_join(basicskills_final, comp_spent,
264                   by=c("districtid"="districtid", "yr"="yr"))
265 #join with English Learner spending
266 match <- full_join(match, el_spent,
267                   by=c("districtid"="districtid", "yr"="yr"))
268
269
270 #add districtname and other info about the district
271 #filter out district without a name (MENTOR - no money either)
272 compare_districts <- left_join(match, district_list %>% select(id_number,
district_name, county, metro7county, location), by=c("districtid"="id_number")) %>%
273   filter(district_name!='NA')
274
275
276
277
278
279 #fill in null values
280 compare_districts$total_compensatory_revenue[is.na(compare_districts$total_compensatory
_revenue)] <- 0
281 compare_districts$el_revenue[is.na(compare_districts$el_revenue)] <- 0
282 compare_districts$el_concentration_revenue[is.na(compare_districts$el_concentration_rev
enue)] <- 0
283 compare_districts$comp_spent[is.na(compare_districts$comp_spent)] <- 0
284 compare_districts$el_spent[is.na(compare_districts$el_spent)] <- 0
285
286
287 #add fields
288
289 #calculate total spent
290 #different between spending and revenue
291 #pct spent is the percentage of revenue that was spent
292 compare_districts <- compare_districts %>%
293   mutate(tot_spent = comp_spent+el_spent,
294          diff=round(total_basic_skills_revenue-tot_spent,2),
295          pctspent = (tot_spent/total_basic_skills_revenue)*100)
296
297
298
299
300 #add a column that puts that pct spending into buckets
301 compare_districts <- compare_districts %>%
302   mutate(scope = case_when(pctspent==0 ~ 'none',
303                             pctspent==100~'100%',
304                             pctspent>100 ~ 'over spent',

```

```

305         pctspent<100 ~'under spent',
306         TRUE ~'check'))
307
308     #calculate the difference in spending as a percentage of revenue
309     #then put them into buckets
310     #anything that is 15% or more over or under are potentially problematic (according to
311     MDE)
312     compare_districts <- compare_districts %>% mutate(
313     diffpct=round((diff/total_basic_skills_revenue)*100,1),
314     diffscope =
315     case_when(diffpct>=14.49~'over by 15% or more',
316             diffpct<14.49
317             & diffpct>9.49~'over by 10%-14%',
318             diffpct<=9.49
319             & diffpct>0 ~'over by less than 10%',
320             diffpct==0
321             ~'even',
322             diffpct<0 &
323             diffpct> -9.49 ~'under by less than 10%',
324             diffpct>
325             -14.49 & diffpct< -9.49~'under by 10-14%',
326             diffpct<=
327             -14.49~'under by 15% or more',
328             TRUE~'something
329             went wrong'))
330
331     #Count up how many fell in each bucket in 2018
332     compare_districts %>% filter(yr==2018) %>% group_by(diffscope) %>% summarise(count=n())
333
334     write.csv(compare_districts, './output/districts_basicskills_totals.csv',
335     row.names=FALSE)
336
337
338
339
340
341
342
343

```