



```

      name: <unnamed>
      log: /hdir/0/jhaber/Projects/charter_data/stats_team/logs/data_preparation_101
> 019.smcl
      log type: smcl
      opened on: 10 Oct 2019, 17:56:58

```

```

1 .
2 . * Import data:
3 . import delimited data/charters_stats_2015_v2a_counts_inqnew.csv, clear
   (69 vars, 6,947 obs)

4 .
5 .
6 . ** -----
7 . ** CREATE LAGGED TEST SCORE VARIABLES: 2013-14
8 . ** -----
9 .
10. * Create variables to store accuracy level of scores data and all scores data:
11. gen readlevel13 = .
   (6,947 missing values generated)

12. gen mathlevel13 = .
   (6,947 missing values generated)

13. gen readall13 = .
   (6,947 missing values generated)

14. gen mathall13 = .
   (6,947 missing values generated)

15. gen readingscore = reading1314
   (1,896 missing values generated)

16. gen mathscore = math1314
   (2,157 missing values generated)

17.
18. * Fix >=99 value:
19. replace readingscore = "99" if readingscore == "GE99"
   (36 real changes made)

20. replace mathscore = "99" if mathscore == "GE99"
   (11 real changes made)

21.
22. * Create var for percentile scores:
23. gen readpct = real(readingscore)
   (5,186 missing values generated)

24. gen mathpct = real(mathscore)
   (5,327 missing values generated)

25. * Merge with readall13/mathall13:
26. replace readall13 = readpct if readpct != .
   (1,761 real changes made)

27. replace mathall13 = mathpct if mathpct != .
   (1,620 real changes made)

```

```
28. * Record accuracy level of score:
29. replace readlevel13 = 1 if readpct != .
   (1,761 real changes made)

30. replace mathlevel13 = 1 if mathpct != .
   (1,620 real changes made)

31.
32. * Create var for quintile scores and record level, merge with readall13/mathall13:
33. gen read5 = .
   (6,947 missing values generated)

34. gen math5 = .
   (6,947 missing values generated)

35.
36. replace read5 = 2 if readingscore == "LE5"
   (8 real changes made)

37. replace read5 = 7 if readingscore == "6-9"
   (13 real changes made)

38. replace read5 = 12 if readingscore == "10-14"
   (38 real changes made)

39. replace read5 = 17 if readingscore == "15-19"
   (49 real changes made)

40. replace read5 = 22 if readingscore == "20-24"
   (55 real changes made)

41. replace read5 = 27 if readingscore == "25-29"
   (68 real changes made)

42. replace read5 = 32 if readingscore == "30-34"
   (89 real changes made)

43. replace read5 = 37 if readingscore == "35-39"
   (78 real changes made)

44. replace read5 = 42 if readingscore == "40-44"
   (108 real changes made)

45. replace read5 = 47 if readingscore == "45-49"
   (106 real changes made)

46. replace read5 = 52 if readingscore == "50-54"
   (132 real changes made)

47. replace read5 = 57 if readingscore == "55-59"
   (127 real changes made)

48. replace read5 = 62 if readingscore == "60-64"
   (136 real changes made)

49. replace read5 = 67 if readingscore == "65-69"
   (148 real changes made)

50. replace read5 = 72 if readingscore == "70-74"
   (144 real changes made)
```

```
51. replace read5 = 77 if readingscore == "75-79"
    (153 real changes made)
52. replace read5 = 82 if readingscore == "80-84"
    (121 real changes made)
53. replace read5 = 87 if readingscore == "85-89"
    (137 real changes made)
54. replace read5 = 92 if readingscore == "90-94"
    (102 real changes made)
55. replace read5 = 97 if readingscore == "GE95"
    (82 real changes made)
56.
57. replace math5 = 2 if mathscore == "LE5"
    (40 real changes made)
58. replace math5 = 7 if mathscore == "6-9"
    (36 real changes made)
59. replace math5 = 12 if mathscore == "10-14"
    (49 real changes made)
60. replace math5 = 17 if mathscore == "15-19"
    (77 real changes made)
61. replace math5 = 22 if mathscore == "20-24"
    (66 real changes made)
62. replace math5 = 27 if mathscore == "25-29"
    (68 real changes made)
63. replace math5 = 32 if mathscore == "30-34"
    (82 real changes made)
64. replace math5 = 37 if mathscore == "35-39"
    (110 real changes made)
65. replace math5 = 42 if mathscore == "40-44"
    (107 real changes made)
66. replace math5 = 44 if mathscore == "45-49"
    (114 real changes made)
67. replace math5 = 52 if mathscore == "50-54"
    (140 real changes made)
68. replace math5 = 57 if mathscore == "55-59"
    (139 real changes made)
69. replace math5 = 62 if mathscore == "60-64"
    (131 real changes made)
70. replace math5 = 67 if mathscore == "65-69"
    (125 real changes made)
71. replace math5 = 72 if mathscore == "70-74"
    (116 real changes made)
```

```
72. replace math5 = 77 if mathscore == "75-79"
    (103 real changes made)

73. replace math5 = 82 if mathscore == "80-84"
    (98 real changes made)

74. replace math5 = 87 if mathscore == "85-89"
    (80 real changes made)

75. replace math5 = 92 if mathscore == "90-94"
    (53 real changes made)

76. replace math5 = 97 if mathscore == "GE95"
    (46 real changes made)

77.
78. replace readall13 = read5 if read5 != .
    (1,894 real changes made)

79. replace mathall13 = math5 if math5 != .
    (1,780 real changes made)

80.
81. replace readlevel13 = 5 if read5 != .
    (1,894 real changes made)

82. replace mathlevel13 = 5 if math5 != .
    (1,780 real changes made)

83.
84. * Create var for decile scores and record level:
85. gen read10 = .
    (6,947 missing values generated)

86. gen math10 = .
    (6,947 missing values generated)

87.
88. replace read10 = 5 if readingscore == "LE10"
    (23 real changes made)

89. replace read10 = 15 if readingscore == "11-19"
    (40 real changes made)

90. replace read10 = 25 if readingscore == "20-29"
    (49 real changes made)

91. replace read10 = 35 if readingscore == "30-39"
    (55 real changes made)

92. replace read10 = 45 if readingscore == "40-49"
    (78 real changes made)

93. replace read10 = 55 if readingscore == "50-59"
    (84 real changes made)

94. replace read10 = 65 if readingscore == "60-69"
    (83 real changes made)

95. replace read10 = 75 if readingscore == "70-79"
    (88 real changes made)
```

```

96. replace read10 = 85 if readingscore == "80-89"
   (54 real changes made)

97. replace read10 = 95 if readingscore == "GE90"
   (53 real changes made)

98.
99. replace math10 = 5 if mathscore == "LE10"
   (68 real changes made)

100 replace math10 = 15 if mathscore == "11-19"
    (53 real changes made)

101 replace math10 = 25 if mathscore == "20-29"
    (69 real changes made)

102 replace math10 = 35 if mathscore == "30-39"
    (68 real changes made)

103 replace math10 = 45 if mathscore == "40-49"
    (74 real changes made)

104 replace math10 = 55 if mathscore == "50-59"
    (76 real changes made)

105 replace math10 = 65 if mathscore == "60-69"
    (59 real changes made)

106 replace math10 = 75 if mathscore == "70-79"
    (53 real changes made)

107 replace math10 = 85 if mathscore == "80-89"
    (48 real changes made)

108 replace math10 = 95 if mathscore == "GE90"
    (28 real changes made)

109
110 replace readall13 = read10 if read10 != .
    (607 real changes made)

111 replace mathall13 = math10 if math10 != .
    (596 real changes made)

112
113 replace readlevel13 = 10 if read10 != .
    (607 real changes made)

114 replace mathlevel13 = 10 if math10 != .
    (596 real changes made)

115
116 * Create var for ventile scores and record level:
117 gen read20 = .
    (6,947 missing values generated)

118 gen math20 = .
    (6,947 missing values generated)

119 replace read20 = 10 if readingscore == "LE20"
    (53 real changes made)

```

```
120 replace read20 = 30 if readingscore == "21-39"
    (60 real changes made)
121 replace read20 = 50 if readingscore == "40-59"
    (98 real changes made)
122 replace read20 = 70 if readingscore == "60-79"
    (91 real changes made)
123 replace read20 = 90 if readingscore == "GE80"
    (75 real changes made)
124
125 replace math20 = 10 if mathscore == "LE20"
    (89 real changes made)
126 replace math20 = 30 if mathscore == "21-39"
    (86 real changes made)
127 replace math20 = 50 if mathscore == "40-59"
    (102 real changes made)
128 replace math20 = 70 if mathscore == "60-79"
    (72 real changes made)
129 replace math20 = 90 if mathscore == "GE80"
    (31 real changes made)
130
131 replace readall13 = read20 if read20 != .
    (377 real changes made)
132 replace mathall13 = math20 if math20 != .
    (380 real changes made)
133
134 replace readlevel13 = 20 if read20 != .
    (377 real changes made)
135 replace mathlevel13 = 20 if math20 != .
    (380 real changes made)
136
137 * Create var for median scores and record level:
138 gen read50 = .
    (6,947 missing values generated)
139 gen math50 = .
    (6,947 missing values generated)
140
141 replace read50 = 25 if readingscore == "LT50"
    (97 real changes made)
142 replace read50 = 75 if readingscore == "GE50"
    (132 real changes made)
143
144 replace math50 = 25 if mathscore == "LT50"
    (154 real changes made)
```

```

145 replace math50 = 75 if mathscore == "GE50"
    (77 real changes made)

146
147 replace readall13 = read50 if read50 != .
    (229 real changes made)

148 replace mathall13 = math50 if math50 != .
    (231 real changes made)

149
150 replace readlevel13 = 50 if read50 != .
    (229 real changes made)

151 replace mathlevel13 = 50 if math50 != .
    (231 real changes made)

152
153 drop readingscore mathscore readpct mathpct read5 math5 read10 math10 read20 math20
    > read50 math50

154
155
156 ** -----
157 ** CREATE LAGGED TEST SCORE VARIABLES: 2014-15
158 ** -----
159
160 * Create variables to store accuracy level of scores data and all scores data:
161 gen readlevel14 = .
    (6,947 missing values generated)

162 gen mathlevel14 = .
    (6,947 missing values generated)

163 gen readall14 = .
    (6,947 missing values generated)

164 gen mathall14 = .
    (6,947 missing values generated)

165 gen readingscore = reading1415
    (941 missing values generated)

166 gen mathscore = math1415
    (937 missing values generated)

167
168 * Fix >=99 value:
169 replace readingscore = "99" if readingscore == "GE99"
    (5 real changes made)

170 replace mathscore = "99" if mathscore == "GE99"
    (2 real changes made)

171
172 * Create var for percentile scores:
173 gen readpct = real(readingscore)
    (4,537 missing values generated)

174 gen mathpct = real(mathscore)
    (4,612 missing values generated)

```

```
175 * Merge with readall14/mathall14:
176 replace readall14 = readpct if readpct != .
    (2,410 real changes made)

177 replace mathall14 = mathpct if mathpct != .
    (2,335 real changes made)

178 * Record accuracy level of score:
179 replace readlevel14 = 1 if readpct != .
    (2,410 real changes made)

180 replace mathlevel14 = 1 if mathpct != .
    (2,335 real changes made)

181
182 * Create var for quintile scores and record level, merge with readall14/mathall14:
183 gen read5 = .
    (6,947 missing values generated)

184 gen math5 = .
    (6,947 missing values generated)

185
186 replace read5 = 2 if readingscore == "LE5"
    (38 real changes made)

187 replace read5 = 7 if readingscore == "6-9"
    (57 real changes made)

188 replace read5 = 12 if readingscore == "10-14"
    (90 real changes made)

189 replace read5 = 17 if readingscore == "15-19"
    (128 real changes made)

190 replace read5 = 22 if readingscore == "20-24"
    (153 real changes made)

191 replace read5 = 27 if readingscore == "25-29"
    (153 real changes made)

192 replace read5 = 32 if readingscore == "30-34"
    (159 real changes made)

193 replace read5 = 37 if readingscore == "35-39"
    (157 real changes made)

194 replace read5 = 42 if readingscore == "40-44"
    (154 real changes made)

195 replace read5 = 47 if readingscore == "45-49"
    (149 real changes made)

196 replace read5 = 52 if readingscore == "50-54"
    (177 real changes made)

197 replace read5 = 57 if readingscore == "55-59"
    (158 real changes made)

198 replace read5 = 62 if readingscore == "60-64"
    (162 real changes made)
```



```
199 replace read5 = 67 if readingscore == "65-69"
    (162 real changes made)
200 replace read5 = 72 if readingscore == "70-74"
    (88 real changes made)
201 replace read5 = 77 if readingscore == "75-79"
    (97 real changes made)
202 replace read5 = 82 if readingscore == "80-84"
    (81 real changes made)
203 replace read5 = 87 if readingscore == "85-89"
    (55 real changes made)
204 replace read5 = 92 if readingscore == "90-94"
    (33 real changes made)
205 replace read5 = 97 if readingscore == "GE95"
    (33 real changes made)
206
207 replace math5 = 2 if mathscore == "LE5"
    (134 real changes made)
208 replace math5 = 7 if mathscore == "6-9"
    (112 real changes made)
209 replace math5 = 12 if mathscore == "10-14"
    (154 real changes made)
210 replace math5 = 17 if mathscore == "15-19"
    (157 real changes made)
211 replace math5 = 22 if mathscore == "20-24"
    (191 real changes made)
212 replace math5 = 27 if mathscore == "25-29"
    (198 real changes made)
213 replace math5 = 32 if mathscore == "30-34"
    (166 real changes made)
214 replace math5 = 37 if mathscore == "35-39"
    (166 real changes made)
215 replace math5 = 42 if mathscore == "40-44"
    (156 real changes made)
216 replace math5 = 44 if mathscore == "45-49"
    (136 real changes made)
217 replace math5 = 52 if mathscore == "50-54"
    (113 real changes made)
218 replace math5 = 57 if mathscore == "55-59"
    (127 real changes made)
219 replace math5 = 62 if mathscore == "60-64"
    (99 real changes made)
```

```
220 replace math5 = 67 if mathscore == "65-69"
    (93 real changes made)

221 replace math5 = 72 if mathscore == "70-74"
    (66 real changes made)

222 replace math5 = 77 if mathscore == "75-79"
    (60 real changes made)

223 replace math5 = 82 if mathscore == "80-84"
    (44 real changes made)

224 replace math5 = 87 if mathscore == "85-89"
    (24 real changes made)

225 replace math5 = 92 if mathscore == "90-94"
    (26 real changes made)

226 replace math5 = 97 if mathscore == "GE95"
    (38 real changes made)

227
228 replace readall14 = read5 if read5 != .
    (2,284 real changes made)

229 replace mathall14 = math5 if math5 != .
    (2,260 real changes made)

230
231 replace readlevel14 = 5 if read5 != .
    (2,284 real changes made)

232 replace mathlevel14 = 5 if math5 != .
    (2,260 real changes made)

233
234 * Create var for decile scores and record level:
235 gen read10 = .
    (6,947 missing values generated)

236 gen math10 = .
    (6,947 missing values generated)

237
238 replace read10 = 5 if readingscore == "LE10"
    (62 real changes made)

239 replace read10 = 15 if readingscore == "11-19"
    (48 real changes made)

240 replace read10 = 25 if readingscore == "20-29"
    (71 real changes made)

241 replace read10 = 35 if readingscore == "30-39"
    (93 real changes made)

242 replace read10 = 45 if readingscore == "40-49"
    (102 real changes made)

243 replace read10 = 55 if readingscore == "50-59"
    (86 real changes made)
```

```
244 replace read10 = 65 if readingscore == "60-69"
    (72 real changes made)

245 replace read10 = 75 if readingscore == "70-79"
    (62 real changes made)

246 replace read10 = 85 if readingscore == "80-89"
    (41 real changes made)

247 replace read10 = 95 if readingscore == "GE90"
    (32 real changes made)

248
249 replace math10 = 5 if mathscore == "LE10"
    (127 real changes made)

250 replace math10 = 15 if mathscore == "11-19"
    (91 real changes made)

251 replace math10 = 25 if mathscore == "20-29"
    (110 real changes made)

252 replace math10 = 35 if mathscore == "30-39"
    (88 real changes made)

253 replace math10 = 45 if mathscore == "40-49"
    (75 real changes made)

254 replace math10 = 55 if mathscore == "50-59"
    (53 real changes made)

255 replace math10 = 65 if mathscore == "60-69"
    (50 real changes made)

256 replace math10 = 75 if mathscore == "70-79"
    (33 real changes made)

257 replace math10 = 85 if mathscore == "80-89"
    (30 real changes made)

258 replace math10 = 95 if mathscore == "GE90"
    (35 real changes made)

259
260 replace readall14 = read10 if read10 != .
    (669 real changes made)

261 replace mathall14 = math10 if math10 != .
    (692 real changes made)

262
263 replace readlevel14 = 10 if read10 != .
    (669 real changes made)

264 replace mathlevel14 = 10 if math10 != .
    (692 real changes made)

265
266 * Create var for ventile scores and record level:
267 gen read20 = .
    (6,947 missing values generated)
```

```
268 gen math20 = .  
    (6,947 missing values generated)  
269 replace read20 = 10 if readingscore == "LE20"  
    (90 real changes made)  
270 replace read20 = 30 if readingscore == "21-39"  
    (76 real changes made)  
271 replace read20 = 50 if readingscore == "40-59"  
    (83 real changes made)  
272 replace read20 = 70 if readingscore == "60-79"  
    (64 real changes made)  
273 replace read20 = 90 if readingscore == "GE80"  
    (36 real changes made)  
274  
275 replace math20 = 10 if mathscore == "LE20"  
    (169 real changes made)  
276 replace math20 = 30 if mathscore == "21-39"  
    (79 real changes made)  
277 replace math20 = 50 if mathscore == "40-59"  
    (77 real changes made)  
278 replace math20 = 70 if mathscore == "60-79"  
    (32 real changes made)  
279 replace math20 = 90 if mathscore == "GE80"  
    (30 real changes made)  
280  
281 replace readall14 = read20 if read20 != .  
    (349 real changes made)  
282 replace mathall14 = math20 if math20 != .  
    (387 real changes made)  
283  
284 replace readlevel14 = 20 if read20 != .  
    (349 real changes made)  
285 replace mathlevel14 = 20 if math20 != .  
    (387 real changes made)  
286  
287 * Create var for median scores and record level:  
288 gen read50 = .  
    (6,947 missing values generated)  
289 gen math50 = .  
    (6,947 missing values generated)  
290  
291 replace read50 = 25 if readingscore == "LT50"  
    (136 real changes made)
```

```

292 replace read50 = 75 if readingscore == "GE50"
    (79 real changes made)

293
294 replace math50 = 25 if mathscore == "LT50"
    (179 real changes made)

295 replace math50 = 75 if mathscore == "GE50"
    (57 real changes made)

296
297 replace readall14 = read50 if read50 != .
    (215 real changes made)

298 replace mathall14 = math50 if math50 != .
    (236 real changes made)

299
300 replace readlevel14 = 50 if read50 != .
    (215 real changes made)

301 replace mathlevel14 = 50 if math50 != .
    (236 real changes made)

302
303 drop readingscore mathscore readpct mathpct read5 math5 read10 math10 read20 math20
    > read50 math50

304
305
306 ** -----
307 ** RECODE TEST SCORE VARIABLES: 2015-16
308 ** -----
309
310 * Create variables to store accuracy level of scores data and all scores data:
311 gen readlevel15 = .
    (6,947 missing values generated)

312 gen mathlevel15 = .
    (6,947 missing values generated)

313 gen readall15 = .
    (6,947 missing values generated)

314 gen mathall15 = .
    (6,947 missing values generated)

315 gen readingscore = reading1516
    (649 missing values generated)

316 gen mathscore = math1516
    (651 missing values generated)

317
318 * Fix >=99 value:
319 replace readingscore = "99" if readingscore == "GE99"
    (8 real changes made)

320 replace mathscore = "99" if mathscore == "GE99"
    (4 real changes made)

```

```
321
322 * Create var for percentile scores:
323 gen readpct = real(readingscore)
      (4,349 missing values generated)

324 gen mathpct = real(mathscore)
      (4,414 missing values generated)

325 * Merge with readall15/mathall15:
326 replace readall15 = readpct if readpct != .
      (2,598 real changes made)

327 replace mathall15 = mathpct if mathpct != .
      (2,533 real changes made)

328 * Record accuracy level of score:
329 replace readlevel15 = 1 if readpct != .
      (2,598 real changes made)

330 replace mathlevel15 = 1 if mathpct != .
      (2,533 real changes made)

331
332 * Create var for quintile scores and record level, merge with readall15/mathall15:
333 gen read5 = .
      (6,947 missing values generated)

334 gen math5 = .
      (6,947 missing values generated)

335
336 replace read5 = 2 if readingscore == "LE5"
      (49 real changes made)

337 replace read5 = 7 if readingscore == "6-9"
      (48 real changes made)

338 replace read5 = 12 if readingscore == "10-14"
      (102 real changes made)

339 replace read5 = 17 if readingscore == "15-19"
      (121 real changes made)

340 replace read5 = 22 if readingscore == "20-24"
      (151 real changes made)

341 replace read5 = 27 if readingscore == "25-29"
      (168 real changes made)

342 replace read5 = 32 if readingscore == "30-34"
      (170 real changes made)

343 replace read5 = 37 if readingscore == "35-39"
      (162 real changes made)

344 replace read5 = 42 if readingscore == "40-44"
      (158 real changes made)

345 replace read5 = 47 if readingscore == "45-49"
      (144 real changes made)
```

```
346 replace read5 = 52 if readingscore == "50-54"
    (162 real changes made)
347 replace read5 = 57 if readingscore == "55-59"
    (167 real changes made)
348 replace read5 = 62 if readingscore == "60-64"
    (144 real changes made)
349 replace read5 = 67 if readingscore == "65-69"
    (130 real changes made)
350 replace read5 = 72 if readingscore == "70-74"
    (108 real changes made)
351 replace read5 = 77 if readingscore == "75-79"
    (114 real changes made)
352 replace read5 = 82 if readingscore == "80-84"
    (71 real changes made)
353 replace read5 = 87 if readingscore == "85-89"
    (59 real changes made)
354 replace read5 = 92 if readingscore == "90-94"
    (50 real changes made)
355 replace read5 = 97 if readingscore == "GE95"
    (30 real changes made)
356
357 replace math5 = 2 if mathscore == "LE5"
    (146 real changes made)
358 replace math5 = 7 if mathscore == "6-9"
    (108 real changes made)
359 replace math5 = 12 if mathscore == "10-14"
    (156 real changes made)
360 replace math5 = 17 if mathscore == "15-19"
    (166 real changes made)
361 replace math5 = 22 if mathscore == "20-24"
    (192 real changes made)
362 replace math5 = 27 if mathscore == "25-29"
    (184 real changes made)
363 replace math5 = 32 if mathscore == "30-34"
    (157 real changes made)
364 replace math5 = 37 if mathscore == "35-39"
    (135 real changes made)
365 replace math5 = 42 if mathscore == "40-44"
    (161 real changes made)
366 replace math5 = 44 if mathscore == "45-49"
    (171 real changes made)
```

```
367 replace math5 = 52 if mathscore == "50-54"
    (119 real changes made)
368 replace math5 = 57 if mathscore == "55-59"
    (119 real changes made)
369 replace math5 = 62 if mathscore == "60-64"
    (83 real changes made)
370 replace math5 = 67 if mathscore == "65-69"
    (84 real changes made)
371 replace math5 = 72 if mathscore == "70-74"
    (83 real changes made)
372 replace math5 = 77 if mathscore == "75-79"
    (57 real changes made)
373 replace math5 = 82 if mathscore == "80-84"
    (56 real changes made)
374 replace math5 = 87 if mathscore == "85-89"
    (39 real changes made)
375 replace math5 = 92 if mathscore == "90-94"
    (34 real changes made)
376 replace math5 = 97 if mathscore == "GE95"
    (29 real changes made)
377
378 replace readall15 = read5 if read5 != .
    (2,308 real changes made)
379 replace mathall15 = math5 if math5 != .
    (2,279 real changes made)
380
381 replace readlevel15 = 5 if read5 != .
    (2,308 real changes made)
382 replace mathlevel15 = 5 if math5 != .
    (2,279 real changes made)
383
384 * Create var for decile scores and record level:
385 gen read10 = .
    (6,947 missing values generated)
386 gen math10 = .
    (6,947 missing values generated)
387
388 replace read10 = 5 if readingscore == "LE10"
    (73 real changes made)
389 replace read10 = 15 if readingscore == "11-19"
    (74 real changes made)
390 replace read10 = 25 if readingscore == "20-29"
    (105 real changes made)
```



```
391 replace read10 = 35 if readingscore == "30-39"
    (69 real changes made)
392 replace read10 = 45 if readingscore == "40-49"
    (89 real changes made)
393 replace read10 = 55 if readingscore == "50-59"
    (77 real changes made)
394 replace read10 = 65 if readingscore == "60-69"
    (46 real changes made)
395 replace read10 = 75 if readingscore == "70-79"
    (43 real changes made)
396 replace read10 = 85 if readingscore == "80-89"
    (38 real changes made)
397 replace read10 = 95 if readingscore == "GE90"
    (33 real changes made)
398
399 replace math10 = 5 if mathscore == "LE10"
    (169 real changes made)
400 replace math10 = 15 if mathscore == "11-19"
    (90 real changes made)
401 replace math10 = 25 if mathscore == "20-29"
    (92 real changes made)
402 replace math10 = 35 if mathscore == "30-39"
    (77 real changes made)
403 replace math10 = 45 if mathscore == "40-49"
    (62 real changes made)
404 replace math10 = 55 if mathscore == "50-59"
    (53 real changes made)
405 replace math10 = 65 if mathscore == "60-69"
    (31 real changes made)
406 replace math10 = 75 if mathscore == "70-79"
    (32 real changes made)
407 replace math10 = 85 if mathscore == "80-89"
    (21 real changes made)
408 replace math10 = 95 if mathscore == "GE90"
    (41 real changes made)
409
410 replace readall15 = read10 if read10 != .
    (647 real changes made)
411 replace mathall15 = math10 if math10 != .
    (668 real changes made)
412
```

```
413 replace readlevel15 = 10 if read10 != .
    (647 real changes made)

414 replace mathlevel15 = 10 if math10 != .
    (668 real changes made)

415
416 * Create var for ventile scores and record level:
417 gen read20 = .
    (6,947 missing values generated)

418 gen math20 = .
    (6,947 missing values generated)

419 replace read20 = 10 if readingscore == "LE20"
    (152 real changes made)

420 replace read20 = 30 if readingscore == "21-39"
    (94 real changes made)

421 replace read20 = 50 if readingscore == "40-59"
    (69 real changes made)

422 replace read20 = 70 if readingscore == "60-79"
    (39 real changes made)

423 replace read20 = 90 if readingscore == "GE80"
    (24 real changes made)

424
425 replace math20 = 10 if mathscore == "LE20"
    (227 real changes made)

426 replace math20 = 30 if mathscore == "21-39"
    (79 real changes made)

427 replace math20 = 50 if mathscore == "40-59"
    (42 real changes made)

428 replace math20 = 70 if mathscore == "60-79"
    (36 real changes made)

429 replace math20 = 90 if mathscore == "GE80"
    (19 real changes made)

430
431 replace readall15 = read20 if read20 != .
    (378 real changes made)

432 replace mathall15 = math20 if math20 != .
    (403 real changes made)

433
434 replace readlevel15 = 20 if read20 != .
    (378 real changes made)

435 replace mathlevel15 = 20 if math20 != .
    (403 real changes made)

436
```

```

437 * Create var for median scores and record level:
438 gen read50 = .
    (6,947 missing values generated)

439 gen math50 = .
    (6,947 missing values generated)

440
441 replace read50 = 25 if readingscore == "LT50"
    (177 real changes made)

442 replace read50 = 75 if readingscore == "GE50"
    (54 real changes made)

443
444 replace math50 = 25 if mathscore == "LT50"
    (217 real changes made)

445 replace math50 = 75 if mathscore == "GE50"
    (47 real changes made)

446
447 replace readall15 = read50 if read50 != .
    (231 real changes made)

448 replace mathall15 = math50 if math50 != .
    (264 real changes made)

449
450 replace readlevel15 = 50 if read50 != .
    (231 real changes made)

451 replace mathlevel15 = 50 if math50 != .
    (264 real changes made)

452
453
454 ** -----
455 ** CREATE AND MODIFY VARIABLES
456 ** -----
457
458 * Use "None" to cluster missing CMOs:
459 replace cmoname = "None" if missing(cmoname)
    (4,582 real changes made)

460
461 * Drop if missing level variables or inquiry/discipline variables:
462 drop if missing(primary) | missing(middle) | missing(high) | missing(otherlevel)
    (19 observations deleted)

463 drop if missing(inquiry_full_log) | missing(numwords) | numwords < 10
    (1,119 observations deleted)

464
465 * Replace with missing any remaining odd codes:
466 replace pocschoolcount = . if pocschoolcount== -12 | pocschoolcount== -54
    (6 real changes made, 6 to missing)

467 drop if missing(pocschoolcount)
    (13 observations deleted)

```

```

468
469 * Convert CMO and state variables into numeric with labels:
470 encode cmoname, gen(cmonum)

471 encode state, gen(statenum)

472
473 * Convert number PDFs to a percentage:
474 gen pctpdfs = numpdfs/numpages

475
476 * Revised teacher count to later create student/teacher ratio:
477 gen teachers = teachersccd if teachersccd != .
    (1,693 missing values generated)

478 replace teachers = teacherscrdc if teachersccd == . & teacherscrdc != .
    (1,541 real changes made)

479
480 * Create log-scaled vars to account for time-shifted effects:
481 gen lnage = ln(age)

482 gen lnstudents = ln(students)
    (12 missing values generated)

483 gen lnteachers = ln(teachers)
    (201 missing values generated)

484
485 * Rescale SD & academic variables so on same scale as dichotomous (0/1) variables:
486 replace pocsd = pocsd/100
    (5,767 real changes made)

487 replace povertysd = povertysd/100
    (5,765 real changes made)

488 replace readall13 = readall13/100
    (4,037 real changes made)

489 replace mathall13 = mathall13/100
    (3,830 real changes made)

490 replace readall14 = readall14/100
    (4,960 real changes made)

491 replace mathall14 = mathall14/100
    (4,948 real changes made)

492 replace readall15 = readall15/100
    (5,158 real changes made)

493 replace mathall15 = mathall15/100
    (5,142 real changes made)

494
495
496 ** -----
497 ** FINALIZE & SAVE DATA
498 ** -----
499

```

```

500 * Label variables:
501 labvars inquiry_seed_count "Emphasis: IBL seed (#)" inquiry_seed_prop "Emphasis: IBL
> seed (%)" inquiry_seed_log "Emphasis: IBL seed (logged ratio)" ///
> inquiry_narrow_count "Emphasis: IBL narrow (#)" inquiry_narrow_prop "Emphasis: IBL n
> arrow (%)" inquiry_narrow_log "Emphasis: IBL narrow (logged ratio)" ///
> inquiry_full_count "Emphasis: IBL full (#)" inquiry_full_prop "Emphasis: IBL full (%
> )" inquiry_full_log "Emphasis: IBL full (logged ratio)" ///
> inquiry_full_nohands_count "Emphasis: IBL full w/o hands-on (#)" inquiry_full_nohand
> s_prop "Emphasis: IBL full w/o hands-on (%)" inquiry_full_nohands_log "Emphasis: IBL
> full w/o hands-on (logged ratio)" ///
> pocschool "% students of color (school)" pocschoolcount "# students of color" ///
> povertyschool "% students in poverty (school)" povertyschoolcount "# students in pov
> erty" ///
> pocsd "% people of color (school district)" povertysd "% people in poverty (school d
> istrict)" ///
> cmoname "CMO" state "State" geodistrict "School district" ///
> readall13 "% proficiency in RLA 2013-14" readall14 "% proficiency in RLA 2014-15" re
> adall15 "% proficiency in RLA 2015-16" ///
> mathall13 "% proficiency in math 2013-14" mathall14 "% proficiency in math 2014-15"
> mathall15 "% proficiency in math 2015-16" ///
> readlevel13 "RLA blurring 2013-14" readlevel14 "RLA blurring 2014-15" readlevel15 "R
> LA blurring 2015-16" ///
> mathlevel13 "Math blurring 2013-14" mathlevel14 "Math blurring 2014-15" mathlevel15
> "Math blurring 2015-16" ///
> primary "Primary school (binary)" middle "Primary school (binary)" high "Primary sch
> ool (binary)" otherlevel "Primary school (binary)" ///
> age "Years open" lnage "Years open (log)" students "# students" lnstudents "# studen
> ts (log)" teachers "# teachers" lnteachers "# teachers (log)" ///
> urban "Urban locale" pctpdfs "% PDF pages" numwords "# words", ///
> alternate

502
503
504 * Save data to disk, indicate not yet imputed:
505 cd "/hdir/0/jhaber/Projects/charter_data/sorting-schools-2019/"
/hdir/0/jhaber/Projects/charter_data/sorting-schools-2019

506 save "data/charter_schools_data_no_imputations.dta", replace
file data/charter_schools_data_no_imputations.dta saved

507 export delimited using "data/charter_schools_data_no_imputations.csv", replace
file data/charter_schools_data_no_imputations.csv saved

508
509 log close
name: <unnamed>
log: /hdir/0/jhaber/Projects/charter_data/stats_team/logs/data_preparation_101
> 019.smcl
log type: smcl
closed on: 10 Oct 2019, 17:57:04

```
