04_Summary_Document_pdf

Tom

May 9, 2018

Goals for project

Calculate percentage of students whose **STAAR reading scores** are:

- Approaching grade level
  - Overall
  - By school type (Elementary, middle, high school)
  - By gender
  - By race
- Meets grade level
  - Overall
  - By school type (E, M, S)
  - By gender
  - By race
- Create at least one graph

Summary of Wrangling

I recoded negative values (-1 and -3) as well as "." as NA, since I assumed these were missing or otherwise incorrect data. I excluded gradetype == "B", as I was unsure what it was in relation to. The main issue with the data was it was in a non-tidy wide format, so I `gathered` the data to convert to tall format. I then extracted out the various codes for race, group, gender, and gradetype as individual factors rather than a single combined factor. I also labeled these factors with more descriptive (eg CB00 = African American student). I `filtered` to only include Reading performance, and then `joined` the data back together into either meeting or approaching datasets. I converted all the student values (numerator, denominator, rate) to numbers and checked to make sure that Rate was equivalent to percentage. I then saved both an Excel and RDS output of the clean data for storage or use in summary calculations/graphing.

Table 1: STAAR Reading level by Group, Grade-type, and Year

Grade-type	Group	Year	Meeting Grade Level (%)	Approaching Grade Level (%)	Number of Districts	Total Students taking STAAR
Elementary	All Students	2016	36.2%	67.8%	150	34967
Elementary	All Students	2017	39.8%	68%	151	35263
Elementary	Female	2016	39.6%	72.1%	150	17053
Elementary	Female	2017	43.4%	71.8%	151	17042
Elementary	Male	2016	33%	63.8%	150	17914
Elementary	Male	2017	36.4%	64.4%	151	18221
Elementary	Econ Disadv	2016	35.1%	67%	150	31245
Elementary	Econ Disadv	2017	38.5%	66.9%	151	31717
Elementary	Special Ed	2016	32.7%	51.7%	93	1999
Elementary	Special Ed	2017	44.7%	51.7%	103	2262
Elementary	At Risk	2016	29.2%	61.4%	150	26023
Elementary	At Risk	2017	32.8%	62.4%	151	23446
Elementary	ELL	2016	37.8%	68.2%	146	17873
Elementary	ELL	2017	41.5%	70.2%	147	17964
Middle-school	All Students	2016	33.8%	66.1%	42	28692
Middle-school	All Students	2017	36.7%	67.4%	42	28710
Middle-school	Female	2016	35.4%	69.2%	41	13881
Middle-school	Female	2017	38.3%	71%	40	13882
Middle-school	Male	2016	29.9%	61.6%	41	14811
Middle-school	Male	2017	33.4%	63.1%	40	14827
Middle-school	Econ Disadv	2016	32.3%	65.3%	42	25317
Middle-school	Econ Disadv	2017	35.3%	67.1%	42	25757
Middle-school	Special Ed	2016	28%	32.9%	31	2537
Middle-school	Special Ed	2017	23.6%	33.6%	33	2529
Middle-school	At Risk	2016	24.8%	59.4%	42	21383
Middle-school	At Risk	2017	26.8%	60.9%	42	21466
Middle-school	ELL	2016	26.7%	61.6%	42	13488
Middle-school	ELL	2017	29.3%	63.2%	42	14110
Secondary	All Students	2016	50.7%	67.1%	37	28396
Secondary	All Students	2017	55.4%	65.2%	37	28649
Secondary	Female	2016	54%	70.4%	36	13642
Secondary	Female	2017	57.2%	67.5%	36	13584
Secondary	Male	2016	45%	62.4%	36	14754
Secondary	Male	2017	51.5%	61.5%	36	15065
Secondary	Econ Disadv	2016	49.6%	66.6%	37	23817
Secondary	Econ Disadv	2017	54.3%	64.7%	37	24384
Secondary	Special Ed	2016	13.3%	23.4%	22	2696
Secondary	Special Ed	2017	20.9%	21.6%	22	2938
Secondary	At Risk	2016	33.9%	58.3%	35	21054
Secondary	At Risk	2017	42.8%	58.4%	37	21187
Secondary	ELL	2016	31.3%	52%	34	9650
Secondary	ELL	2017	36.2%	51.8%	33	10901

Summary Tables

I focused on splitting the data into race and groups (gender, special categories). I am presenting them here as those two tables but I also saved an additional table with ALL the data combined.

Table 2: STAAR Reading level by Race, Grade-type, and Year

Grade-type	Group	Year	Meeting Grade Level (%)	Approaching Grade Level (%)	Number of Districts	Total Students taking STAAR
Elementary	African American	2016	28%	61.7%	104	6631
Elementary	African American	2017	32.8%	59.9%	107	6989
Elementary	White	2016	67.1%	88.3%	40	1403
Elementary	White	2017	67.9%	86.1%	34	1345
Elementary	Hispanic	2016	38.2%	69.3%	140	24601
Elementary	Hispanic	2017	42.5%	71.1%	136	24370
Elementary	Two or More Races	2016	85%	95%	2	15
Elementary	Two or More Races	2017	80.3%	97.5%	5	45
Elementary	Asian	2016	77.1%	85.2%	11	196
Elementary	Asian	2017	69.9%	83.3%	12	246
Middle-school	African American	2016	34.8%	62.7%	39	5910
Middle-school	African American	2017	36.4%	65.2%	40	5863
Middle-school	White	2016	69.5%	81.1%	24	1092
Middle-school	White	2017	66%	81.4%	22	1115
Middle-school	Hispanic	2016	33.5%	66.8%	42	20965
Middle-school	Hispanic	2017	35.9%	67.7%	42	21066
Middle-school	Two or More Races	2016	85.6%	92.6%	6	56
Middle-school	Two or More Races	2017	77.9%	89.6%	9	84
Middle-school	Asian	2016	62.4%	79.3%	12	270
Middle-school	Asian	2017	75.5%	86.9%	13	259
Secondary	African American	2016	48.7%	65.1%	35	6385
Secondary	African American	2017	50.7%	59.4%	35	6426
Secondary	White	2016	79.7%	80.4%	20	993
Secondary	White	2017	79.1%	73.1%	18	1006
Secondary	Hispanic	2016	50.4%	67.2%	35	19993
Secondary	Hispanic	2017	55.8%	66.2%	37	20469
Secondary	Two or More Races	2016	88.2%	98%	3	41
Secondary	Two or More Races	2017	92.9%	100%	1	15
Secondary	Asian	2016	73.6%	78.9%	11	419
Secondary	Asian	2017	80.1%	79.8%	10	367

Graphs

I wanted to focus on examining the data by race, and specifically for the 4 most common groups (African american, asian, hispanic, and white). I plotted two graphs, one for meeting reading standards and one for approaching reading standards.

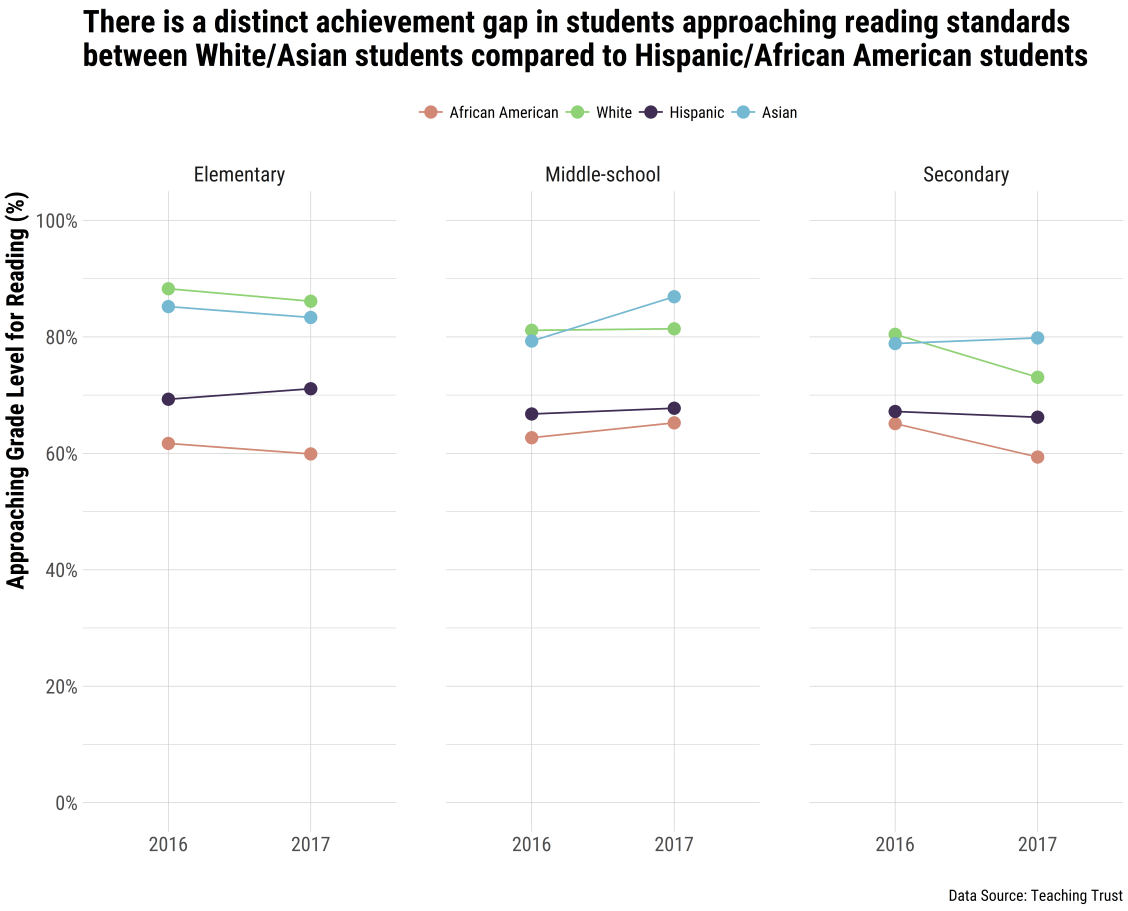

Figure 1: Fig. 1 2016-17 STAAR Reading performance approaching grade level standards by race

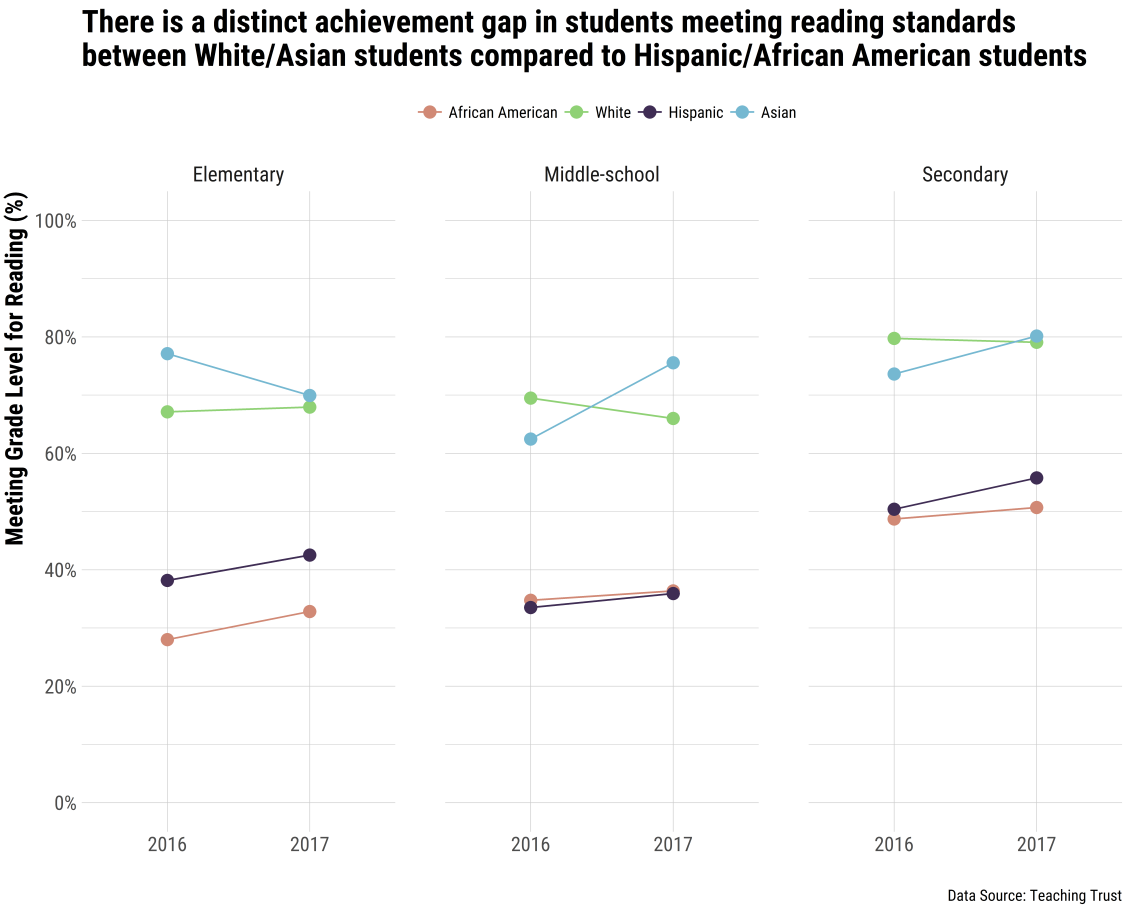

Figure 2: Fig. 2 2016-17 STAAR Reading performance meeting grade level standards by race

One thing I learned

I really haven't had to do that many summary tables for my PhD, as we typically graph everything. So I wanted to end up with nice looking tables in a .rmd report as something to learn. I focused on the `kableExtra` package which is intended to create beautiful tables quickly and painlessly. I also appreciate the fact that there was a consistent coding scheme for the data, and although it is not how data is typically arranged in my world, it was done consistently and clearly!

Additionally, at the beginning of the assignment I spent a lot of time digging through the coding system to make sure I understood how to break it up and not have to hardcode all the various combinations of groups/races/subject/year. I'm really pleased with the pseudo-database I formed that has the various groups assigned, as it saved me a lot of coding.

How did I fill gap?

I really spent longer than expected making sure I understood some of the ways I would need to alter the raw data to appear nice in a table, so I dug through the documentation for `kableExtra` on GitHub as seen here. I wanted to end up with a useful, clean, and attractive table, and this made it a lot easier. This required me renaming some of the variables so they had nice descriptive titles rather than simple snake-case titles.

60-90 Min adequate?

I definitely went over in time, I could have completed the assignment inside 90 minutes, but I spent an extra hour at the beginning planning out HOW I wanted to start the project, what I wanted to achieve and the most efficient way to do that. I also spent some extra time going back through the assignment and really making sure I felt comfortable passing it off to a colleague for them to understand what I was trying to do. All in it took me probably 2.5 hours to get through everything, not including time to read through the `kableExtra` documentation further.

I really enjoyed working through this project, and I like that y'all provided take-home assignments!