what

batman

Compiled on Sun May 7 12:21:39 2017.

*Source file* ⇒ get\_star\_results.Rmd

get data from tab delimited files and csvs

# store a list of files  
files <- dir("star\_results/", recursive = T)  
excludeFiles <- grepl("entities|Entities", files, perl = T)  
files <- files[!excludeFiles]  
files <- lapply(files, function(x) {paste("star\_results/", x, sep = "")})  
  
# store 1998 names (34 cols) so we can make later tables consistent  
names1998 <- c("District\_Code",  
 "Test\_Date",  
 "Record\_Type",  
 "Summary\_Type",  
 "Reading\_Mean\_SS",  
 "Reading\_Mean\_PR",  
 "Reading\_PAC75",  
 "Reading\_PACAT50",  
 "Reading\_PAC25",  
 "Math\_Mean\_SS",  
 "Math\_Mean\_PR",  
 "Math\_PAC75",  
 "Math\_PACAT50",  
 "Math\_PAC25",  
 "Language\_Mean\_SS",  
 "Language\_Mean\_PR",  
 "Language\_PAC75",  
 "Language\_PACAT50",  
 "Language\_PAC25",  
 "Spelling\_Mean\_SS",  
 "Spelling\_Mean\_PR",  
 "Spelling\_PAC75",  
 "Spelling\_PACAT50",  
 "Spelling\_PAC25",  
 "Science\_Mean\_SS",  
 "Science\_Mean\_PR",  
 "Science\_PAC75",  
 "Science\_PACAT50",  
 "Science\_PAC25",  
 "Social\_Science\_Mean\_SS",  
 "Social\_Science\_Mean\_PR",  
 "Social\_Science\_PAC75",  
 "Social\_Science\_PACAT50",  
 "Social\_Science\_PAC25"  
 )  
  
# function to group by District, Record and Summary types  
# then calculate the means of each column  
get\_means98 <- function(t) {  
 t[, -1] <- t[, -1] %>%  
 sapply(as.numeric)  
  
 t <- t %>%  
 group\_by(District\_Code, Record\_Type, Summary\_Type) %>%  
 summarise\_each(funs(mean(., na.rm = T)))  
  
 return (t)  
}  
  
# assumes that t is a data frame with columns  
# District\_Code, School\_Code, Year, Enrollment, PercentProficient  
get\_means01 <- function(t) {  
 t[, -1] <- t[, -1] %>%  
 sapply(as.numeric)  
  
 colnames(t) <- c("District\_Code",  
 "School\_Code",  
 "Year",  
 "Enrollment",  
 "PercentProficient")  
 t <- t %>%  
 group\_by(District\_Code, School\_Code) %>%  
 summarise\_each(funs(mean(., na.rm = T)))  
  
 return (t)  
}  
  
## now get the stuff we want  
for (f in files) {  
 year\_matches <- stringr::str\_match(f, "star\_results/(\\d{4})/.\*")  
 year <- year\_matches[[2]]  
  
 # eff these years  
 # don't serialize them  
 if (year %in% c("1998", "1999", "2000")) {  
 # 1998  
 if (year == "1998") {  
 cur <- read.table(f, sep = "\t", header = T)  
 cur <- cur %>%  
 select(-c(County\_Code, School\_Code,  
 County\_Name, School\_Name,  
 District\_Name, Grade\_Level,  
 Total\_Enrolled, Total\_Tested,  
 Reading\_Total\_Valid, Math\_Total\_Valid,  
 Language\_Total\_Valid, Spelling\_Total\_Valid,  
 Science\_Total\_Valid, Social\_Science\_Total\_Valid,  
 Filler))  
  
 cur <- get\_means98(cur)  
 }   
 # 1999  
 else if (year == "1999") {  
 cur <- read.table(f, sep = "\t")  
 cur <- cur %>%  
 select(-c(1, 3, 7:9, 10, 16, 22, 28, 34, 40, 46:77))  
  
 # Column names are not listed in the file, so we'll set them here  
 colnames(cur) <- names1998  
 cur <- get\_means98(cur)  
 }  
 # 2000  
 else if (year == "2000") {  
 cur <- read.table(f, sep = "\t", header = T)  
 cur <- cur %>%  
 select(-c(1, 3, 4, 8:10, 11, 17, 23, 29, 35, 41, 47:79))  
 colnames(cur) <- names1998  
 cur <- get\_means98(cur)  
 }  
 } else {  
 # 2001  
 if (year == "2001") {  
 cur <- read.table(f, sep = "\t", header = T)  
 cur <- read.table("star\_results/2001/2001\_tdtestdata.tab", sep = "\t", header = T)  
 cur <- cur %>%  
 select(c(2, 3, 5, 10, 20))  
 cur <- get\_means01(cur)  
 }  
 #2002  
 else if (year == "2002") {  
 cur <- read.csv(f)  
 cur <- cur %>%  
 select(c(2, 3, 5, 10, 20))  
 cur <- get\_means01(cur)  
 }  
 # 2003 - 2012  
 else {  
 cur <- read.csv(f)  
 cur <- cur %>%  
 select(c(2, 3, 5, 15, 19))  
 cur <- get\_means01(cur)  
 }  
  
 # serialize to binary file  
 binary <- paste("star\_result\_binaries/", year, ".feather", sep = "")  
 write\_feather(cur, binary)  
 }  
}

## Warning in lapply(X = X, FUN = FUN, ...): NAs introduced by coercion  
  
## Warning in lapply(X = X, FUN = FUN, ...): NAs introduced by coercion  
  
## Warning in lapply(X = X, FUN = FUN, ...): NAs introduced by coercion  
  
## Warning in lapply(X = X, FUN = FUN, ...): NAs introduced by coercion  
  
## Warning in lapply(X = X, FUN = FUN, ...): NAs introduced by coercion  
  
## Warning in lapply(X = X, FUN = FUN, ...): NAs introduced by coercion  
  
## Warning in lapply(X = X, FUN = FUN, ...): NAs introduced by coercion  
  
## Warning in lapply(X = X, FUN = FUN, ...): NAs introduced by coercion  
  
## Warning in lapply(X = X, FUN = FUN, ...): NAs introduced by coercion  
  
## Warning in lapply(X = X, FUN = FUN, ...): NAs introduced by coercion  
  
## Warning in lapply(X = X, FUN = FUN, ...): NAs introduced by coercion  
  
## Warning in lapply(X = X, FUN = FUN, ...): NAs introduced by coercion  
  
## Warning in lapply(X = X, FUN = FUN, ...): NAs introduced by coercion  
  
## Warning in lapply(X = X, FUN = FUN, ...): NAs introduced by coercion  
  
## Warning in lapply(X = X, FUN = FUN, ...): NAs introduced by coercion  
  
## Warning in lapply(X = X, FUN = FUN, ...): NAs introduced by coercion  
  
## Warning in lapply(X = X, FUN = FUN, ...): NAs introduced by coercion  
  
## Warning in lapply(X = X, FUN = FUN, ...): NAs introduced by coercion  
  
## Warning in lapply(X = X, FUN = FUN, ...): NAs introduced by coercion  
  
## Warning in lapply(X = X, FUN = FUN, ...): NAs introduced by coercion  
  
## Warning in lapply(X = X, FUN = FUN, ...): NAs introduced by coercion  
  
## Warning in lapply(X = X, FUN = FUN, ...): NAs introduced by coercion  
  
## Warning in lapply(X = X, FUN = FUN, ...): NAs introduced by coercion  
  
## Warning in lapply(X = X, FUN = FUN, ...): NAs introduced by coercion  
  
## Warning in lapply(X = X, FUN = FUN, ...): NAs introduced by coercion  
  
## Warning in lapply(X = X, FUN = FUN, ...): NAs introduced by coercion  
  
## Warning in lapply(X = X, FUN = FUN, ...): NAs introduced by coercion  
  
## Warning in lapply(X = X, FUN = FUN, ...): NAs introduced by coercion  
  
## Warning in lapply(X = X, FUN = FUN, ...): NAs introduced by coercion  
  
## Warning in lapply(X = X, FUN = FUN, ...): NAs introduced by coercion

## Warning in scan(file = file, what = what, sep = sep, quote = quote, dec =  
## dec, : embedded nul(s) found in input

## Warning in lapply(X = X, FUN = FUN, ...): NAs introduced by coercion  
  
## Warning in lapply(X = X, FUN = FUN, ...): NAs introduced by coercion  
  
## Warning in lapply(X = X, FUN = FUN, ...): NAs introduced by coercion  
  
## Warning in lapply(X = X, FUN = FUN, ...): NAs introduced by coercion  
  
## Warning in lapply(X = X, FUN = FUN, ...): NAs introduced by coercion  
  
## Warning in lapply(X = X, FUN = FUN, ...): NAs introduced by coercion  
  
## Warning in lapply(X = X, FUN = FUN, ...): NAs introduced by coercion  
  
## Warning in lapply(X = X, FUN = FUN, ...): NAs introduced by coercion  
  
## Warning in lapply(X = X, FUN = FUN, ...): NAs introduced by coercion  
  
## Warning in lapply(X = X, FUN = FUN, ...): NAs introduced by coercion

# TEST STUFF FOR SOME YEAR  
res2001 <- read\_feather("star\_result\_binaries/2001.feather")  
head(res2001)

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| District\_Code | School\_Code | Year | Enrollment | PercentProficient |
| 0 | 0 | 2001 | 5457.80905079009 | 16.28213430945 |
| 10017 | 0 | 2001 | 41.96064814815 | 2.67857142857 |
| 10017 | 130401 | 2001 | 40.70846394984 | 2.26086956522 |
| 10017 | 130419 | 2001 | 8.72131147541 | 5.22222222222 |
| 10017 | 130427 | 2001 | 10.07692307692 | 3.33333333333 |
| 10033 | 0 | 2001 | 4.75277777778 | 0.00000000000 |

res2002 <- read\_feather("star\_result\_binaries/2002.feather")  
head(res2002)

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| District\_Code | School\_Code | Year | Enrollment | PercentProficient |
| 0 | 0 | 2002 | 4928.72088895101 | 17.09532062392 |
| 10017 | 0 | 2002 | 28.45127118644 | 1.76470588235 |
| 10017 | 130401 | 2002 | 23.23218997362 | 1.76923076923 |
| 10017 | 130419 | 2002 | 6.50364963504 | 4.00000000000 |
| 10017 | 130427 | 2002 | 11.14509803922 | 0.68000000000 |
| 10033 | 0 | 2002 | 6.69838056680 | 3.13043478261 |

res2003 <- read\_feather("star\_result\_binaries/2003.feather")  
head(res2003)

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| District\_Code | School\_Code | Year | Enrollment | PercentProficient |
| 0 | 0 | 2003 | 8675.04839149400 | 20.748663101604 |
| 10017 | 0 | 2003 | 41.43396226415 | 1.210526315789 |
| 10017 | 130401 | 2003 | 30.44897959184 | 0.142857142857 |
| 10017 | 130419 | 2003 | 7.11111111111 | 1.500000000000 |
| 10017 | 130427 | 2003 | 11.63636363636 | 5.200000000000 |
| 10033 | 0 | 2003 | 12.65882352941 | 1.875000000000 |

res2004 <- read\_feather("star\_result\_binaries/2004.feather")  
head(res2004)

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| District\_Code | School\_Code | Year | Enrollment | PercentProficient |
| 0 | 0 | 2004 | 9141.8037983331 | 18.24951644101 |
| 10017 | 0 | 2004 | 53.7826086957 | 2.07142857143 |
| 10017 | 130401 | 2004 | 39.0937500000 | 1.40000000000 |
| 10017 | 130419 | 2004 | 18.4444444444 | 3.58333333333 |
| 10017 | 130427 | 2004 | 11.9090909091 | 1.20000000000 |
| 10025 | 0 | 2004 | 1.0000000000 | NaN |

res2005 <- read\_feather("star\_result\_binaries/2005.feather")  
head(res2005)

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| District\_Code | School\_Code | Year | Enrollment | PercentProficient |
| 0 | 0 | 2005 | 7091.0325493885 | 19.54177831912 |
| 10017 | 0 | 2005 | 30.4324324324 | 2.36842105263 |
| 10017 | 130401 | 2005 | 24.8333333333 | 1.50000000000 |
| 10017 | 130419 | 2005 | 12.2647058824 | 6.61538461539 |
| 10017 | 130427 | 2005 | 11.9090909091 | 1.80000000000 |
| 10025 | 0 | 2005 | 1.0000000000 | NaN |

res2006 <- read\_feather("star\_result\_binaries/2006.feather")  
head(res2006)

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| District\_Code | School\_Code | Year | Enrollment | PercentProficient |
| 0 | 0 | 2006 | 7429.39910564561 | 19.68806899856 |
| 10017 | 0 | 2006 | 28.89473684211 | 2.25000000000 |
| 10017 | 109835 | 2006 | 37.82539682540 | 17.76315789474 |
| 10017 | 130401 | 2006 | 18.64285714286 | 3.00000000000 |
| 10017 | 130419 | 2006 | 10.27777777778 | 1.78571428571 |
| 10017 | 130427 | 2006 | 9.36363636364 | 3.44444444444 |

res2007 <- read\_feather("star\_result\_binaries/2007.feather")  
head(res2007)

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| District\_Code | School\_Code | Year | Enrollment | PercentProficient |
| 0 | 0 | 2007 | 7085.8164916509 | 19.78021978022 |
| 10017 | 0 | 2007 | 27.2325581395 | 2.73333333333 |
| 10017 | 109835 | 2007 | 47.7922077922 | 17.84090909091 |
| 10017 | 112607 | 2007 | 37.0000000000 | 7.00000000000 |
| 10017 | 130401 | 2007 | 19.8620689655 | 3.30000000000 |
| 10017 | 130419 | 2007 | 11.1470588235 | 2.00000000000 |

res2008 <- read\_feather("star\_result\_binaries/2008.feather")  
head(res2008)

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| District\_Code | School\_Code | Year | Enrollment | PercentProficient |
| 0 | 0 | 2008 | 6190.2503384986 | 20.60331339162 |
| 10017 | 0 | 2008 | 33.7941176471 | 3.94117647059 |
| 10017 | 109835 | 2008 | 45.4302325581 | 19.34042553191 |
| 10017 | 112607 | 2008 | 41.5000000000 | 3.46666666667 |
| 10017 | 130401 | 2008 | 28.8947368421 | 4.66666666667 |
| 10017 | 130419 | 2008 | 14.5161290323 | 2.85714285714 |

res2009 <- read\_feather("star\_result\_binaries/2009.feather")  
head(res2009)

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| District\_Code | School\_Code | Year | Enrollment | PercentProficient |
| 0 | 0 | 2009 | 4843.8735813820 | 20.85123801917 |
| 10017 | 0 | 2009 | 39.6333333333 | 3.50000000000 |
| 10017 | 109835 | 2009 | 39.6506024096 | 20.38297872340 |
| 10017 | 112607 | 2009 | 35.0384615385 | 8.38888888889 |
| 10017 | 118489 | 2009 | 30.5000000000 | 34.68421052632 |
| 10017 | 130401 | 2009 | 27.3333333333 | 2.20000000000 |

res2010 <- read\_feather("star\_result\_binaries/2010.feather")  
head(res2010)

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| District\_Code | School\_Code | Year | Enrollment | PercentProficient |
| 0 | 0 | 2010 | 4625.5711731536 | 21.73630831643 |
| 10017 | 0 | 2010 | 25.2162162162 | 2.81818181818 |
| 10017 | 109835 | 2010 | 41.4500000000 | 20.60465116279 |
| 10017 | 112607 | 2010 | 37.2424242424 | 9.00000000000 |
| 10017 | 118489 | 2010 | 31.7142857143 | 36.82608695652 |
| 10017 | 130401 | 2010 | 12.3030303030 | 1.00000000000 |

res2011 <- read\_feather("star\_result\_binaries/2011.feather")  
head(res2011)

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| District\_Code | School\_Code | Year | Enrollment | PercentProficient |
| 0 | 0 | 2011 | 4463.3398831652 | 21.71097156811 |
| 10017 | 0 | 2011 | 23.5952380952 | 6.15000000000 |
| 10017 | 109835 | 2011 | 42.1707317073 | 22.19148936170 |
| 10017 | 112607 | 2011 | 38.4594594595 | 11.28571428571 |
| 10017 | 118489 | 2011 | 28.4722222222 | 30.05000000000 |
| 10017 | 130401 | 2011 | 17.4230769231 | 2.54545454546 |

res2012 <- read\_feather("star\_result\_binaries/2012.feather")  
head(res2012)

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| District\_Code | School\_Code | Year | Enrollment | PercentProficient |
| 0 | 0 | 2012 | 4461.2023460411 | 21.65053763441 |
| 10017 | 0 | 2012 | 22.5952380952 | 2.86666666667 |
| 10017 | 109835 | 2012 | 38.6136363636 | 20.76086956522 |
| 10017 | 112607 | 2012 | 35.1190476190 | 12.30000000000 |
| 10017 | 118489 | 2012 | 27.3695652174 | 35.64516129032 |
| 10017 | 123968 | 2012 | 15.5000000000 | 34.50000000000 |