Generate Star Test Averages vs State Funding Graphs

William Jow

Compiled on Sun May 7 12:25:05 2017.

*Source file* ⇒ mk\_star\_funding\_graphs.Rmd

# scores  
star2001 <- read\_feather("star\_result\_binaries/2001.feather")  
head(star2001)

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| 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 |

star2002 <- read\_feather("star\_result\_binaries/2002.feather")  
head(star2002)

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| 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 |

star2003 <- read\_feather("star\_result\_binaries/2003.feather")  
head(star2003)

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| 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 |

star2004 <- read\_feather("star\_result\_binaries/2004.feather")  
head(star2004)

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| 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 |

star2005 <- read\_feather("star\_result\_binaries/2005.feather")  
head(star2005)

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| 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 |

star2006 <- read\_feather("star\_result\_binaries/2006.feather")  
head(star2006)

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| 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 |

star2007 <- read\_feather("star\_result\_binaries/2007.feather")  
head(star2007)

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| 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 |

star2008 <- read\_feather("star\_result\_binaries/2008.feather")  
head(star2008)

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| 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 |

star2009 <- read\_feather("star\_result\_binaries/2009.feather")  
head(star2009)

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| 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 |

star2010 <- read\_feather("star\_result\_binaries/2010.feather")  
head(star2010)

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| 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 |

star2011 <- read\_feather("star\_result\_binaries/2011.feather")  
head(star2011)

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| 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 |

star2012 <- read\_feather("star\_result\_binaries/2012.feather")  
head(star2012)

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| 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 |

star\_all <- rbind(star2001,  
 star2002,  
 star2003,  
 star2004,  
 star2005,  
 star2006,  
 star2007,  
 star2008,  
 star2009,  
 star2010,  
 star2011,  
 star2012)  
  
# get weighted means and serialize  
star\_weighted <- ddply(star\_all, .(District\_Code, Year), summarise,  
 weighted.mean(PercentProficient, Enrollment, na.rm = T))  
colnames(star\_weighted)[3] <- "PercentProficient"  
star\_weighted$Year <- factor(star\_weighted$Year)  
write\_feather(star\_weighted,  
 "star\_result\_binaries/all\_years\_proficient\_weighted.feather")  
head(star\_weighted)

|  |  |  |
| --- | --- | --- |
| District\_Code | Year | PercentProficient |
| 0 | 2001 | 16.2821343094 |
| 0 | 2002 | 17.0953206239 |
| 0 | 2003 | 20.7486631016 |
| 0 | 2004 | 18.2495164410 |
| 0 | 2005 | 19.5417783191 |
| 0 | 2006 | 19.6880689986 |

# funding  
funding <- read.table("funding1998\_2012.txt", sep = "\t", header = T)  
funding <- select(funding, c(1, 2, 3, 6, 8))  
colnames(funding) <- c("County\_Code", "District", "District\_Code",  
 "Expense\_per\_ADA", "Year")  
funding$Year <- factor(funding$Year)  
head(funding)

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| County\_Code | District | District\_Code | Expense\_per\_ADA | Year |
| 1 | Albany City Unified | 61127 | 5775.28 | 1998 |
| 1 | Berkeley Unified | 61143 | 7117.97 | 1998 |
| 1 | Castro Valley Unified | 61150 | 5075.40 | 1998 |
| 1 | Dublin Unified | 75093 | 5896.15 | 1998 |
| 1 | Emery Unified | 61168 | 6357.31 | 1998 |
| 1 | Fremont Unified | 61176 | 4939.71 | 1998 |

# both for graph stuff  
star\_vs\_funding <- merge(star\_weighted, funding,  
 by = c("District\_Code", "Year"))  
head(star\_vs\_funding)

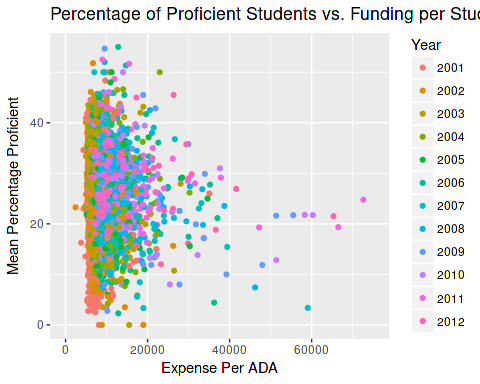
|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| District\_Code | Year | PercentProficient | County\_Code | District | Expense\_per\_ADA |
| 61119 | 2002 | 24.6861916870 | 1 | Alameda City Unified | 6650.32 |
| 61119 | 2004 | 26.3686442256 | 1 | Alameda City Unified | 7255.26 |
| 61119 | 2005 | 28.5264143373 | 1 | Alameda City Unified | 7594.38 |
| 61119 | 2007 | 28.4466812681 | 1 | Alameda City Unified | 8007.97 |
| 61119 | 2008 | 28.1656050565 | 1 | Alameda City Unified | 8640.96 |
| 61119 | 2009 | 27.6284139532 | 1 | Alameda City Unified | 8629.52 |

star\_vs\_funding\_year <- star\_vs\_funding %>%   
 ddply(.(Year),  
 summarise,  
 PercentProficient = mean(PercentProficient, na.rm = T),  
 Expense\_per\_ADA = mean(Expense\_per\_ADA, na.rm = T))  
head(star\_vs\_funding\_year)

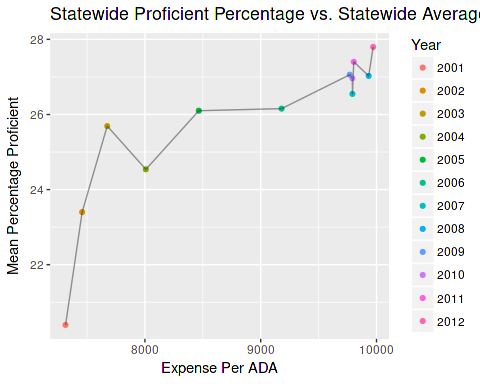
|  |  |  |
| --- | --- | --- |
| Year | PercentProficient | Expense\_per\_ADA |
| 2001 | 20.3956989838 | 7314.54399177 |
| 2002 | 23.3994868370 | 7457.68927909 |
| 2003 | 25.6922111130 | 7674.40737113 |
| 2004 | 24.5385695556 | 8007.90383661 |
| 2005 | 26.1002751050 | 8465.47109731 |
| 2006 | 26.1557228179 | 9180.14126687 |

# graph stuff  
star\_vs\_funding\_plot <- star\_vs\_funding %>% ggplot(aes(x = Expense\_per\_ADA, y = PercentProficient)) +  
 geom\_point(aes(col = Year)) +  
 scale\_x\_continuous(limits = c(0, 75000)) +  
 scale\_y\_continuous(limits = c(0, 55)) +  
 labs(x = "Expense Per ADA",  
 y = "Mean Percentage Proficient",  
 title = "Percentage of Proficient Students vs. Funding per Student by District")  
  
star\_vs\_funding\_year\_plot <- star\_vs\_funding\_year %>% ggplot(aes(x = Expense\_per\_ADA, y = PercentProficient)) +  
 geom\_point(aes(col = Year)) +  
 geom\_line(aes(group = 1), col = "grey20", alpha = 0.5) +  
 labs(x = "Expense Per ADA",  
 y = "Mean Percentage Proficient",  
 title = "Statewide Proficient Percentage vs. Statewide Average Student Funding")  
  
star\_vs\_time\_state\_plot <- star\_vs\_funding\_year %>% ggplot(aes(x = Year, y = PercentProficient)) +  
 geom\_point() +  
 geom\_line(aes(group = 1), col = "grey20", alpha = 0.5) +  
 labs(x = "Year",  
 y = "Mean Percentage Proficient",  
 title = "Statewide Proficient Percentage vs. Time")  
  
# show graphs  
star\_vs\_funding\_plot

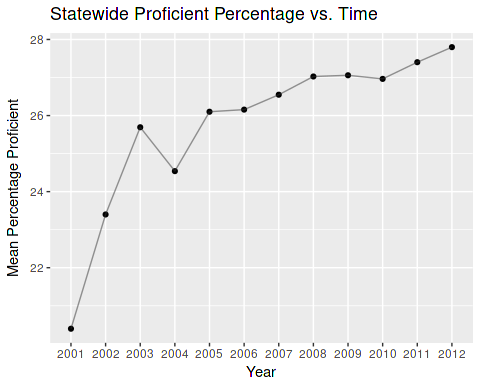
## Warning: Removed 751 rows containing missing values (geom\_point).



star\_vs\_funding\_year\_plot



star\_vs\_time\_state\_plot



# export pdfs of the graphs  
pdf(file = "star\_vs\_funding\_images/star\_vs\_funding\_district.pdf")  
print(star\_vs\_funding\_plot)

## Warning: Removed 751 rows containing missing values (geom\_point).

dev.off()

## png   
## 2

pdf(file = "star\_vs\_funding\_images/star\_vs\_funding\_year.pdf")  
print(star\_vs\_funding\_year\_plot)  
dev.off()

## png   
## 2

pdf(file = "star\_vs\_time\_images/star\_vs\_time\_state.pdf")  
print(star\_vs\_time\_state\_plot)  
dev.off()

## png   
## 2

# do lm stuff  
# "Percentage of Proficient Students vs. Funding per Student by District"  
# Expense\_per\_ADA is the funding per student in the district  
# PercentProficient is the percentage of proficient students in the district  
summary(lm(formula = Expense\_per\_ADA ~ PercentProficient, data = star\_vs\_funding))

##   
## Call:  
## lm(formula = Expense\_per\_ADA ~ PercentProficient, data = star\_vs\_funding)  
##   
## Residuals:  
## Min 1Q Median 3Q Max   
## -6017.08991 -1539.13878 -709.85916 327.72169 169388.96606   
##   
## Coefficients:  
## Estimate Std. Error t value Pr(>|t|)  
## (Intercept) 8755.67912967 127.53362164 68.65389 < 2.22e-16  
## PercentProficient -12.31445517 4.76504305 -2.58433 0.0097698  
##   
## Residual standard error: 3629.25829 on 10763 degrees of freedom  
## (742 observations deleted due to missingness)  
## Multiple R-squared: 0.000620146071, Adjusted R-squared: 0.000527292791   
## F-statistic: 6.67877398 on 1 and 10763 DF, p-value: 0.00976981557

# "Statewide Proficient Percentage vs. Statewide Average Student Funding"  
# Expense\_per\_ADA is the funding per student in the state  
# PercentProficient is the percentage of proficient students in the state  
summary(lm(formula = Expense\_per\_ADA ~ PercentProficient, data = star\_vs\_funding\_year))

##   
## Call:  
## lm(formula = Expense\_per\_ADA ~ PercentProficient, data = star\_vs\_funding\_year)  
##   
## Residuals:  
## Min 1Q Median 3Q Max   
## -1227.665753 -408.928415 159.212112 367.073995 705.720746   
##   
## Coefficients:  
## Estimate Std. Error t value Pr(>|t|)  
## (Intercept) -2221.9757916 2186.8898011 -1.01604 0.33356061  
## PercentProficient 432.9735914 84.6496519 5.11489 0.00045406  
##   
## Residual standard error: 588.458685 on 10 degrees of freedom  
## Multiple R-squared: 0.723467323, Adjusted R-squared: 0.695814055   
## F-statistic: 26.1620916 on 1 and 10 DF, p-value: 0.000454058594

# "Statewide Proficient Percentage vs. Time"  
# Year is... the year  
# PercentProficient is the percentage of proficient students in the state  
summary(lm(formula = as.numeric(Year) ~ PercentProficient, data = star\_vs\_funding\_year))

##   
## Call:  
## lm(formula = as.numeric(Year) ~ PercentProficient, data = star\_vs\_funding\_year)  
##   
## Residuals:  
## Min 1Q Median 3Q Max   
## -3.405409613 -1.047330202 -0.515312566 1.813724438 2.498271329   
##   
## Coefficients:  
## Estimate Std. Error t value Pr(>|t|)  
## (Intercept) -31.373218374 7.292941705 -4.30186 0.00155688  
## PercentProficient 1.470431168 0.282293592 5.20887 0.00039619  
##   
## Residual standard error: 1.96241936 on 10 degrees of freedom  
## Multiple R-squared: 0.730693025, Adjusted R-squared: 0.703762327   
## F-statistic: 27.1323468 on 1 and 10 DF, p-value: 0.000396186076