PSAT Semifinalist

2023-08-06

PSAT Semifinalists in California from 2015-2023

The National Merit Scholarship Program acknowledges the top 0.5% of PSAT scorers as Semifinalists for each state in the United States. There are approximately 16,000 semifinalists each year in the United States.

Different states have different thresholds depending on how well each state does. A list of schools and student names that are Semifinalists are released every year for each state. Therefore, the number of semifinalists for each school district and year can be computed. I did not find the data for 2017 and 2019 online, but they are probably available somewhere.

Since the data is from a PDF, the data is loaded from PDF to Excel by copying and pasting the entire document. Then, some data cleaning is required in R. Some of the school names consist of two lines of text since they are too long, so I manually looked at the school names that require two lines and a had a special case just for those schools with two lines of text.

Trends to notice:

1) The number of semifinalists is very concentrated in certain school districts, especially in suburban areas, likely due to segregation.

2) The number of semifinalists within each school district have remained pretty consistent over the past 8 years.

```
library(readxl)
library(dplyr)
```

```
## Warning: package 'dplyr' was built under R version 4.2.3
```

```
##
## Attaching package: 'dplyr'
```

```
## The following objects are masked from 'package:stats':
##
##     filter, lag
```

```
## The following objects are masked from 'package:base':
##
##     intersect, setdiff, setequal, union
```

```
NationalMerit2023 <- read_excel("C:/Users/willi/Documents/NationalMerit2023.xlsx", col_names = "All") %
  filter(All != "California (continued)" & All != "Semifinalists: 2023 National Merit Scholarship Progra
         & !(All %in% c(0:9)))
NationalMerit2022 <- read_excel("C:/Users/willi/Documents/NationalMerit2022.xlsx", col_names = "All") %
```

```
  filter(All != "California (continued)" & All != "Semifinalists: 2022 National Merit Scholarship Progra
         & !(All %in% c(0:9)))
NationalMerit2021 <- read_excel("C:/Users/willi/Documents/NationalMerit2021.xlsx", col_names = "All") %>
  filter(All != "California (continued)" & All != "Semifinalists: 2021 National Merit Scholarship Progra
         & !(All %in% c(0:9)))
NationalMerit2018 <- read_excel("C:/Users/willi/Documents/NationalMerit2018.xlsx", col_names = "All") %>
  filter(All != "California (continued)" & All != "Semifinalists: 2018 National Merit Scholarship Progra
         & !(All %in% c(0:9)))
NationalMerit2016 <- read_excel("C:/Users/willi/Documents/NationalMerit2016.xlsx", col_names = "All") %>
  filter(All != "California (continued)" & All != "Semifinalists: 2016 National Merit Scholarship Progra
         & !(All %in% c(0:9)))
NationalMerit2015 <- read_excel("C:/Users/willi/Documents/NationalMerit2015.xlsx", col_names = "All") %>
  filter(All != "California (continued)" & All != "Semifinalists: 2015 National Merit Scholarship Progra
         & !(All %in% c(0:9)))
convert_this = function(temp, year){
  vect = temp[[1]]
  blacklist = c("CRYSTAL SPRINGS", "FLINTRIDGE PREPARATORY", "TARBUT V'TORAH COMMUNITY",
                "DR. T. J. OWENS GILROY EARLY", "PACIFIC PALISADES", "SANTA MARGARITA",
                "PALOS VERDES DISTANCE", "UNIVERSITY PREPARATORY", "CALIFORNIA ACADEMY OF",
                "FLINTRIDGE SACRED", "WALDORF SCHOOL OF", "INSPIRE SCHOOL OF ARTS",
                "FAIRMONT PREPARATORY", "SACRED HEART", "QUALIA THE SCHOOL FOR",
                "BELLARMINE COLLEGE", "DEVELOPING VIRTUE", "TEMECULA PREPARATORY",
                "LOS ANGELES CENTER FOR", "BASIS INDEPENDENT", "RUTH ASAWA SCHOOL OF",
                "ACADEMY OF OUR LADY", "JOHN MARSHALL", "ORANGE COUNTY SCHOOL OF",
                "GEORGIANA BRUCE KIRBY", "JEWISH COMMUNITY H. S. OF", "SAN DIEGO H. S. OF",
                "SOUTHERN CALIFORNIA", "GERMAN INTERNATIONAL", "GIRLS ACADEMIC",
                "ST. IGNATIUS COLLEGE", "SACRAMENTO COUNTRY", "GALILEO ACADEMY OF SCIENCE",
                "LOS ANGELES COUNTY H. S. FOR", "FRANCISCO BRAVO MEDICAL", "CIRCLE OF INDEPENDENT")
  temp$School = ""
  temp$City = ""
  temp$Year = year
  school = ""
  city = ""

  skip = FALSE
  for(i in seq_along(vect)){
    if(skip){
      skip = FALSE
      next
    }
    if(vect[i] %in% blacklist){
      school = paste(vect[i], vect[i+1])
      temp$School[i] = NA
      temp$City[i] = NA
      temp$School[i+1] = NA
      temp$City[i+1] = NA
      skip = TRUE
    }
    else if (grepl("H. S.|SCHOOL|ACADEMY|HORIZON PREP|HIGH TECH HIGH|KIPP SAN JOSE COLLEGIATE|FOSHAY LEA
                   vect[i])){
      school = vect[i]
      temp$School[i] = NA
      temp$City[i] = NA
```

```
    }
    else if (grepl("[0-9]", vect[i])){
      temp$School[i] = school
      temp$City[i] = city
    }
    else{
      city = vect[i]
      temp$School[i] = NA
      temp$City[i] = NA
    }
  }
  return(temp)
}

N2023 = convert_this(NationalMerit2023, 2023)
N2022 = convert_this(NationalMerit2022, 2022)
N2021 = convert_this(NationalMerit2021, 2021)
N2018 = convert_this(NationalMerit2018, 2018)
N2016 = convert_this(NationalMerit2016, 2016)
N2015 = convert_this(NationalMerit2015, 2015)

data = bind_rows(N2023, N2022, N2021, N2018, N2016, N2015) %>%
  filter(!is.na(School)) %>%
  tidyr::separate(All, c("LastName","FirstName"), sep = ", ") %>%
  tidyr::separate(LastName, c(NA, "LastName"), sep = " ") %>%
  filter(!is.na(LastName))
```

```
## Warning: Expected 2 pieces. Missing pieces filled with `NA` in 12 rows [1857, 2129,
## 4051, 4279, 4650, 6068, 8091, 8347, 10311, 10549, 12426, 12652].
```

```
## Warning: Expected 2 pieces. Additional pieces discarded in 67 rows [121, 219, 482, 598,
## 724, 850, 882, 1109, 1272, 1281, 1446, 1533, 1972, 2185, 2249, 2299, 2335,
## 2402, 2947, 3022, ...].
```

```
## Warning: Expected 2 pieces. Missing pieces filled with `NA` in 11 rows [1857, 2129,
## 4051, 4279, 6068, 8091, 8347, 10311, 10549, 12426, 12652].
```

```
table_top = data %>%
  group_by(City, Year) %>%
  summarize(Count = n()) %>%
  tidyr::pivot_wider(names_from = "Year", values_from = "Count", names_sort = TRUE) %>%
  arrange(desc(`2023`)) %>%
  ungroup() %>%
  slice_max(order_by = `2023`, n = 25) %>%
  mutate(City = stringr::str_to_title(City))
```

```
## `summarise()` has grouped output by 'City'. You can override using the
## `.groups` argument.
```

```
data$School[data$School == "HARKER SCHOOL"] = "THE HARKER SCHOOL"
data$City[data$City == "LA CANADA"] = "LA CANADA FLINTRIDGE"
```

```r
school_top = data %>%
  group_by(School, Year) %>%
  summarize(Count = n()) %>%
  tidyr::pivot_wider(names_from = "Year", values_from = "Count", names_sort = TRUE) %>%
  arrange(desc(`2023`)) %>%
  ungroup() %>%
  slice_max(order_by = `2023`, n = 20) %>%
  mutate(School = stringr::str_to_title(School))
```

```
## `summarise()` has grouped output by 'School'. You can override using the
## `.groups` argument.
```

```r
get_all = data %>%
  group_by(Year) %>%
  summarize(Count = n())


table_top = table_top %>%
  rbind(c("PERCENT TOTAL", round(unlist(lapply(table_top[2:7], sum, na.rm = TRUE))/get_all[[2]], 3)))

school_top = school_top %>%
  rbind(c("PERCENT TOTAL", round(unlist(lapply(school_top[2:7], sum, na.rm = TRUE))/get_all[[2]], 3)))
```

```
knitr::kable(table_top, caption = "Top Cities by Number of National Merit Semifinalists (PSAT) in Calif
```

Table 1: Top Cities by Number of National Merit Semifinalists (PSAT) in California

City	2015	2016	2018	2021	2022	2023
San Jose	225	244	278	247	238	283
Fremont	161	131	176	110	163	170
San Diego	150	177	171	149	139	146
Cupertino	115	105	99	109	124	103
Irvine	124	88	109	97	91	103
Palo Alto	76	86	76	78	50	95
San Ramon	76	52	59	47	65	70
Pleasanton	55	52	62	41	52	62
San Mateo	8	20	26	33	59	42
Studio City	25	31	20	40	44	40
Los Angeles	22	36	30	38	38	36
Atherton	27	28	33	27	33	35
Mountain View	36	38	56	31	27	34
Saratoga	40	40	43	32	42	32
Dublin	7	8	12	20	41	30
La Jolla	27	25	28	22	31	30
Fullerton	42	64	46	31	36	26
Belmont	6	16	9	13	20	25
La Canada Flintridge	NA	NA	19	18	34	23
Los Altos	10	15	18	14	22	23
Oakland	22	32	27	16	22	23
Danville	15	17	13	13	16	21
Newport Beach	6	8	10	6	18	19
San Francisco	27	36	21	27	12	19
Westlake Village	15	NA	10	6	10	19
PERCENT TOTAL	0.619	0.651	0.648	0.65	0.678	0.693

```
knitr::kable(school_top, caption = "Top Cities by Number of National Merit Semifinalists (PSAT) in Cali
```

Table 2: Top Cities by Number of National Merit Semifinalists (PSAT) in California

School	2015	2016	2018	2021	2022	2023
Mission San Jose H. S.	107	84	91	64	71	81
The Harker School	59	51	47	42	52	67
Dougherty Valley H. S.	55	45	50	40	51	58
Canyon Crest Academy	21	38	52	62	27	48
Henry M. Gunn H. S.	36	46	33	33	15	47
Monta Vista H. S.	68	57	53	63	56	44
American H. S.	15	18	27	13	45	41
Harvard - Westlake School	25	31	20	40	44	40
Lynbrook H. S.	68	48	63	79	69	40
Palo Alto H. S.	25	33	35	37	32	39
Northwood H. S.	30	28	26	27	19	37
Amador Valley H. S.	34	34	46	27	42	36
Homestead H. S.	19	23	20	24	28	33
Irvington H. S.	32	26	49	23	34	33
Basis Independent Silicon Valley	NA	2	21	14	23	31
University H. S.	50	24	34	21	21	30
Saratoga H. S.	38	39	40	31	38	28
Del Norte H. S.	17	25	20	16	17	27
The Nueva School	NA	NA	7	16	25	27
Cupertino H. S.	28	25	25	22	38	26
Foothill H. S.	20	23	17	14	16	26
PERCENT TOTAL	0.351	0.338	0.347	0.364	0.362	0.385