

California's Unlevel Playing Field

Emma Durler, Kylie Scharf, Nicole Yee, Curtis LeQuang

Table of contents

```
library(tidyverse)
library(readxl)
library(writexl)
library(rstatix)
library(tidyverse)
library(kableExtra)
#loading in data
#tuesdata <- tidyuesdayR::tt_load('2022-03-29')
#sports_dat <- tuesdata$sports

#data set creation that doesn't call git hub
sports_dat <- read_excel("data/original_sub/original.xlsx")
```

```
#subsetting and moving data
sports_dat <- sports_dat |>
  filter(sports == "Baseball" | sports == "Softball", state_cd == "CA") |>
  rename(Sport = sports) |>
  group_by(Sport, year) |>
  pivot_wider(id_cols = year:sector_name, names_from = Sport, values_from = exp_men:exp_women) |>
  mutate(diff = exp_men_Baseball - exp_women_Softball) |>
  drop_na(exp_men_Baseball, exp_women_Softball, diff)
```

```
#| fig-cap: "Above entails an analysis of the average differences in expenses from 2015 to 2021"
```

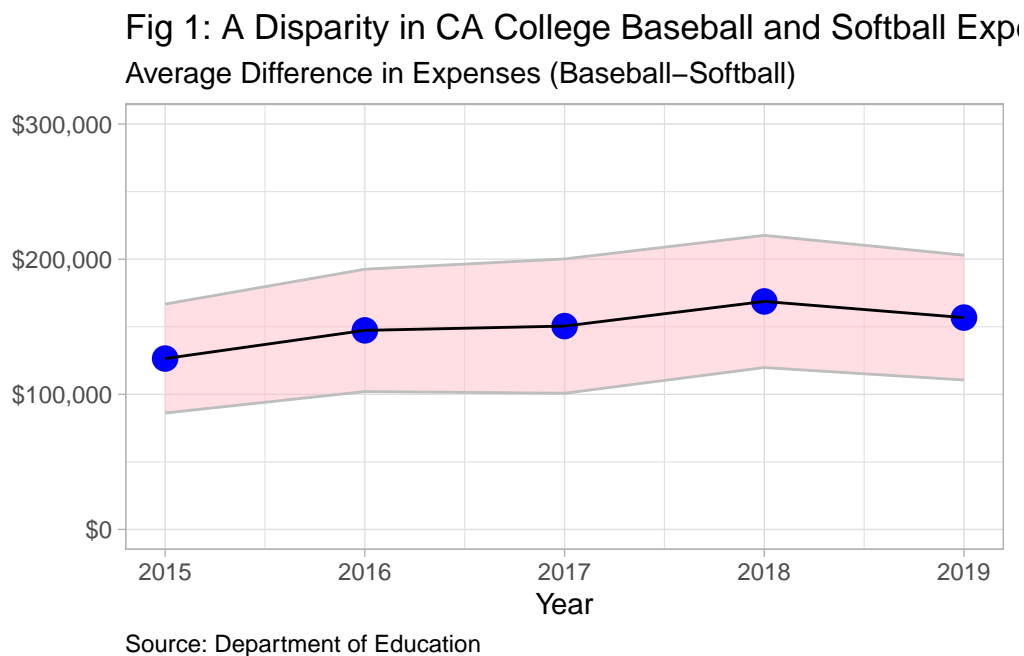
```
plot <- sports_dat |>
  group_by(year) |>
  summarize(mean_exp_men = mean(exp_men_Baseball),
```

```

    mean_exp_women = mean(exp_women_Softball),
    mean_diff = mean(diff),
    stan = sd(diff)/sqrt(n())
) |>
ggplot(aes(y = mean_diff, x = year)) +
  geom_ribbon(aes(ymin = mean_diff - 2*stan, ymax = mean_diff + 2*stan),
            alpha=.5,
            color="grey",
            fill = "pink") +
  theme_bw() +
  theme_light() +
  geom_point(size=4, color="blue") +
  geom_line() +
  labs(x = "Year",
       subtitle = "Average Difference in Expenses (Baseball-Softball)",
       y = "",
       title = "Fig 1: A Disparity in CA College Baseball and Softball Expenses",
       caption = "Source: Department of Education") +
  theme(plot.caption = element_text(hjust = 0)) +
  scale_y_continuous(labels = scales::dollar, limits=c(0,300000))

plot

```



```
#data table using kable
table <- sports_dat |>
  group_by(year) |>
  summarize(mean_exp_men = mean(exp_men_Baseball),
            mean_exp_women = mean(exp_women_Softball),
            mean_diff = mean(diff)
  )

kable(table,
      caption="Average expenses for baseball and softball and the average differences between
      col.names = c("Year",
                    "Baseball Expenses ($)",
                    "Softball Expenses ($)",
                    "Differences ($)")
      ) |>
kable_classic() |>
kable_styling(full_width = FALSE, position = "center") |>
footnote(general = "Source: Department of Education")
```

Table 1: Average expenses for baseball and softball and the average differences between the two for participating California colleges between 2015 and 2019

Year	Baseball Expenses (\$)	Softball Expenses (\$)	Differences (\$)
2015	380821.4	254361.6	126459.8
2016	422943.3	275632.5	147310.8
2017	446679.5	296197.1	150482.4
2018	482359.2	313640.0	168719.2
2019	457521.3	300758.3	156763.0

Note:

Source: Department of Education

```
#analysis
t.test(sports_dat$exp_men_Baseball, sports_dat$exp_women_Softball, paired=TRUE)
```

Paired t-test

data: sports_dat\$exp_men_Baseball and sports_dat\$exp_women_Softball
t = 14.542, df = 658, p-value < 2.2e-16

```
alternative hypothesis: true mean difference is not equal to 0
95 percent confidence interval:
 129676.6 170162.3
sample estimates:
mean difference
 149919.4
```